JOURNAL OF EXPERIMENTAL EDUCATION

A QUARTERLY

Edited by
A. S. BARR

With the Cooperation of

JACOB O. BACH
RAYMOND O. COLLIER
ARTHUR T. JERSILD
H. H. REMMERS
J. WAYNE WRIGHTSTONE

BUREAU OF EDNL. & PSYL.
LIBRARY

September 1960 - June 1961
VOLUME XXIX

DEMBAR PUBLICATIONS, INC.
Madison 3, Wisconsin

TABLE OF CONTENTS

VOLUME XXIX

SEPTEMBER 1960

Children's Spelling Needs and the Implications of Research: Arnie E. Richmond 3

Public Attitudes Toward Education: Donald B. Black et al 23

Scholastic Achievement at Iowa State College Associated with High School Size and Course Pattern: Irvin T. Lathrop 37

A Study of Undergraduate Records of Graduates from Hunter College Who Later Earned Doctorates: Edith B. Chase 49

Syllogistic Reasoning: George N. Beaumariage, Jr. 61

Analysis of Adolescents' Responses to the F Scale: Jack R. Frymier 73

Comparison of Academic Achievement, Aptitudes, and Interest Patterns of Two-Year Technical Students and Four-Year Degree Candidates in Engineering: Louis M. Herman and Martin L. Zeigler 81

An Objective and Empirical Study of the Effects of Laboratory Experience in a Professional Education Course Prior to Student Teaching: Dan Cox 89

A Comparison of Three Ways of Phrasing Likert-Type Attitude Items: Willard E. North and John Schmid 95

Motives in Selecting Elementary and Secondary School Teaching: Gerhard Lang 101

DECEMBER 1960

Children's Perceptions of Their Teachers' Feelings Toward Them Related to Self-Perception, School Achievement and Behavior: Helen H. Davidson and Gerhard Lang 107

Psychological Services in Twenty-Eight Elementary Schools of Columbus, Ohio: Opal E. Parmer 119

Galvanic Skin Responses During Problem Solving: Frank W. Banghart 133

On the Use of Q-Technique in Educational Evaluation and Research: M. Stephen Sheldon and A. Garth Sorenson 143

The Usefulness of a Two-Day Analysis of WISC Sub-Tests in the Diagnosis of Remedial Reading Problems: Lynne Schellberg Hirst 153

Pre-Test and Re-Test Scores in Retention Calculation: Ellis Beecher Little 161

Personality Patterns of Prospective Teachers: Robert F. Peck 169

Factor Analyses of Three Tests of Critical Thinking: Velma I. Rust 177

Prospective Teachers' Estimates of Adolescents' Responses to F Scale Items: Jack R. Frymier 183

Personality Traits Associated with Mathematical Ability—A Plea for Research: Robert H. Riffenburgh 189

Factorial Content of the Iowa Tests of Educational Development and Other Tests: Russell N. Cassell and Edward J. Stancik 193

The House-Drawing Test as a Predictor of First-Grade School Achievement: Harry S. Beck and Robert C. Beck 197

A Device for Quickly Adding Weighted Scores: James W. Creaser 201

MARCH 1961

Teacher Variables of Warmth, Demand and Utilization of Intrinsic Motivation Related to Pupils' Science Interests: A Study Illustrating Several Potentials of Variance-Covariance: Horace B. Reed 205

The Transfer Effect of a Learning Program in Social Causality of an Understanding of Physical Causality: Rolf E. Muuss 231

A Comparison of Scores Obtained by Administering a Test Normally and Visually: H. A. Curtis and R. P. Kropp 249

Nonnormality and Product Moment Correlation: Raymond C. Norris and Howard F. Hjelm 261

Pupil Discovery vs. Direct Instruction: Willis E. Ray 271

Analysis of Questionnaire Responses by Certain Psychologists: Andrew L. Comrey 281

Aptitude and Course Achievement in a Short Missile Technician Course: Harry E. Anderson and Frank Guenter 291

Establishing Criterion Groups for Evaluating Measures of Curiosity: Wallace H. Maw and Ethel W. Maw........................ 299

An Investigation of the Teaching of Chronology in the Sixth Grade: Val E. Arnsdorf.. 307

The Relation Between Arithmetic Research and the Content of Arithmetic Textbooks, 1900-1957: Mother M. Constance Dooley.. 315

A Quick—But Accurate—Approximation to the Standard Deviation of a Distribution: Robert L. Lathrop 319

Comparing Two Methods of Weighting a Set of Scores: Charles E. Hall... 323

JUNE 1961

A Study of the Relationship of Scores on the MMPI to Success in Teaching as Indicated by Supervisory Ratings: Carroll Edward Flanagan 329

A Comparison of Teachers' Response Patterns on the MMPI with Response Patterns of Selected Non-Teacher Groups: Elmer Christian Rieck 355

A Comparison of the Readability of Vocational Agriculture Reference Books with the Reading Ability of the Students Using Them: Ralph R. Bentley and R. Edward Galloway............... 373

The Relationship of Certain Letter Form Variants in the Handwriting of Female Subjects to their Education, I. Q. and Age: Lawrence Epstein, Huntington Hartford, and Irving Tumarkin... 385

A Study of Pathological Doubting as a Response Set: Velma I. Rust... 393

Geometric Aid in Teaching Rectilinear Correlation: John Gaito............... 401

The Journal of Experimental Education

PUBLISHED QUARTERLY

Copyright 1960 by

DEMBAR PUBLICATIONS, INC.
Madison 3, Wisconsin

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 1, September 1960)

CHILDREN'S SPELLING NEEDS AND THE IMPLICATIONS OF RESEARCH

ARNIE E. RICHMOND
Tampa, Florida

ALTHOUGH MANY spelling investigations have been carried out during the past fifty years or more, the results achieved in the teaching of spelling have not been satisfactory when measured by the effectiveness with which children spell in situations that call for written expression. The memorization of a list of isolated words, the principal approach to the learning of spelling in the past, failed to produce efficient spellers. Likewise, the attempts to master the study lists of the present decade are failing to develop adequate spelling ability by a large percentage of children. 1*

Leading authorities are of the opinion that we do have available professional knowledge based on research that would go far toward preventing and solving the spelling problem if this knowledge had been put into practice in the classroom. In view of this situation, the writer was stimulated to conduct a classroom experiment utilizing selected professional knowledge in an attempt to improve the spelling ability of the pupils.

Purpose of This Study

This study was made (a) to determine the spelling needs of the pupils of a sixth grade class through classroom research and experimentation and (b to suggest procedures for spelling instruction to meet those needs by utilizing the implications of scientific investigation and expert opinion.

No attempt is made to present a complete spelling program. It is the opinion of the writer that a spelling program, to be most effective, must evolve from the needs of a particular group of children under the direction and guidance of the teacher. However, basic procedures, considerations, and generalizations applicable to any group in spelling instruction are, the writer believes, inherent in the investigation.

Scope and Background of Study

This study was limited to a suburban consolidated twelve-teacher school serving the community of Landover Hills, Maryland. With the exception of initial data to determine the general spelling competence of the school as a whole in the spontaneous writings of the children, the study was further limited to forty-one pupils of a sixth grade class.

Most of the pupils came from average middle-class homes. The average mental age of the pupils at the beginning of the study in January, as determined by the California Test of Mental Maturity, Intermediate, Short Form, [2] was 145 months with a range from 115 months to 186 months. The I. Q. ranged from 76 to 136 with an average of 104. On the Progressive Achievement Test, Form AA, [3] the average grade placement for reading vocabulary was 5.5, for reading comprehension 6.6, and for language 6.5.

The study extended over a period of five months from January to June. However, basic classroom organization and teaching procedures described later in this section had been established in the preceding months of the school year.

Sources of Data and Procedures

Two chief sources of data were utilized in this study. These were (a) the writings of the pupils in the school, and (b) the professional materials of outstanding educators in the fields of spelling and related areas.

The writings of the pupils were used as sources of data in two ways--first, in determining general spelling competency and in error analysis; and second, for individual word lists. The use of the individual lists will be discussed later in this section.

The Spontaneous Writings in Grades III

---

* Footnotes will be found at the end of the article.

Through VI. As a means of evaluating the general spelling competence of the school as a whole at the beginning of the study, each pupil in Grades III through VI was asked to write a story about any topic of interest, no stipulation being made as to the length of the story. The children were simply asked to spell as well as they could without the help of the teacher. Their spelling errors in these stories were analyzed in terms of all errors made by each grade group, all the errors in the group being pooled.

The Spontaneous Writings of the Sixth Grade Experimental Group. For the sixth grade experimental group, the spelling errors in the January stories were further analyzed to determine:

1. The actual words misspelled for the grade group and for the individual pupils; the repetition of errors for all the pupils; and the most frequently misspelled words.
2. The types of errors made in the spontaneous writings of the children.
3. The occurrence of the misspelled words in (a) The Pupils' Own Vocabulary Spellers,[4] (b) A Basic Life Spelling Vocabulary,[5] and (c) The 2,000 Commonest Words for Spelling.[6]

Again in June, each pupil of the experimental group was asked to write another story. These stories were analyzed as described above and comparisons made with the results of the first stories.

The Sources of Words for Spelling Study. To meet the individual writing needs of the pupils, two main sources of words were used. These were the textbook based on the writings of pupils and the individual pupil lists.

The Spelling Textbook. The Pupils' Own Vocabulary Spellers were selected as the basic textbook because their purposes, methods, and word lists were in agreement with the purposes of this study. The words used in this series were based on the study, A Basic Vocabulary of Elementary School Children, by Henry D. Rinsland.[7]

The complete spelling word list from Grades II through VIII was used as an alphabetical reference list to which the children referred as the need arose.

The Individual Spelling Lists. The pupils were motivated to write freely in a variety of situations. An individual list of the words which each pupil misspelled was kept in a card file at his desk. Whenever a word was misspelled, it was checked with the list from the spelling series referred to above. If the word occurred in this list, it was added to the pupil's file for study. Local words of importance which the child desired to learn were also added to this individual list.

Other Basic Related Pupil Materials. Initial data indicated a need for vocabulary development. As spelling was emphasized at the level of reading, and because of the interrelationship of the language arts, extensive use was made by the pupils of the following materials:

1. Basic Readers: Curriculum Foundation Series.[8]
2. Building Word Power.[9]
3. Thorndike Century Beginning Dictionary.[10]

The Use of Tests and Measurements. The following sources of data for evaluation, diagnosis, and classification and grouping of pupils were used:

1. Temple University Informal Spelling Inventory.[11]
2. Teacher constructed informal spelling inventories using The Pupils' Own Vocabulary Spellers.[12]
3. Gates-Russell Spelling Diagnosis Tests.[13]
4. California Test of Mental Maturity.[14]
5. Progressive Achievement Tests.[15]

All these measurements were given at the beginning of the study, and the first two were repeated, using a different form, at the end of the experiment.

Procedures in Classroom Organization and Instruction. Artley has presented five principles applying to the improvement of spelling ability which have the weight of research behind them, and formed the basis for the classroom organizational and instructional procedures of this study. These are:

1. Each child develops at his own individual rate, and spelling instruction must be differentiated in terms of this rate of development.
2. A spelling vocabulary grows out of the need for expression rather than from a prepared spelling list.
3. Spelling, like any other subject in the curriculum, has to be taught.
4. Spelling has been effectively taught to the extent that the learner has been made independent in his spelling ability.
5. Spelling instruction is most effective when the child has a favorable attitude toward spelling --when he has the desire to spell acceptably.[16]

Needs Revealed by Initial Data. An early examination of data revealed that a wide range of ability existed among the pupils and that their spelling needs were highly individual. However, there was a commonness of needs among three groups within the grade. These were the accelerated who needed enriched and extended experiences in the semantics of words; the so-called average group who needed a maintenance and continuing program in spelling; and the general verbal and linguistic retarded group who needed a program of word study skills applicable to all the language arts. Cutting across groups were pupils who needed specific help in the visual, auditory, speech, kinesthetic, and handwriting skills. All needed an opportunity to write in meaningful situations, and each pupil needed an opportunity to succeed at his level of achievement. Briefly, a functional child development approach to spelling was needed.

A functional approach to spelling required that opportunities be provided for writing in natural situations and that instruction be individualized. These conditions were provided for through the center of interest or functional activity unit in the content areas integrated with a unified language arts program.

The Functional Unit Program. The functional unit or center of interest program in the content areas was characterized by purposeful and satisfying pupil activities--group and individual, a balanced program of intensive and extensive reading, the development of language skills and abilities in social situations, a wise use of many aids to learning, the extension and enrichment of interests, and critical evaluation.

The Unified Language Arts Program. A unified language arts approach to spelling was made on the assumption that the child is integrated and that it is the purpose of the school to foster and maintain that integration. Test data indicated a wide range of language development among the pupils, ranging from the third grade through the eighth grade. This necessitated differentiation of instruction in terms of pupil achievement and needs.

Grouping of pupils was considered a part of an integrated language arts program. Three to four flexible groups were organized in terms of the instructional levels of the pupils. The instructional level of spelling was synchronized with the reading level.

One of the main procedures used for estimating achievement and needs in the group situations was the informal inventory in reading and spelling. These inventories were used for two main purposes. First, they provided a basis for grouping children in terms of their group needs, and second, they served as guides to the needs of the individuals in the group.

Other test data already discussed, and teacher judgment, were used to supplement the informal inventories in the grouping of pupils on levels of achievement. In general, there was little discrepancy among the various evaluation techniques.

Plan of Instruction in the Spelling Period. Daily provision was made for spelling instruction using the textbook and the individual lists. Briefly, the plan of instruction using the textbook was the study-test plan for the retarded pupils and the test-study plan for the others.

Since the pupils were grouped for spelling instruction at the instructional level, the usual daily plan found in spellers of what to do on particular days of the week was disregarded. The group simply proceeded as rapidly as possible in the time allotted, usually fifteen to twenty minutes a day. For example, on any day a group of words could be presented in contextual setting as a reading lesson, since spelling and reading groups were on similar levels, and the skill development exercises connected with those words completed on the same day rather than the following day as suggested by the text.

The learning of the words in the individual list was considered the responsibility of the individual pupils. Provision was made for individual and small group study, usually pairs of similar levels or pairs of different levels, depending upon the activity and pupil choice of partner.

Method of Studying a Word. The studies of Russell[17] and others showed that multiple and complex instructional steps were cumbersome and boring for good spellers and confusing and ineffective for poor spellers. The method employed in this study eliminated all the steps but the vitally useful ones. There were only three steps employed which varied slightly to meet the requirements of longer words. These steps were:

1. Look at the word and say it softly. If it has more than one part (syllable) say it again, part by part, looking at each part as you say it.
2. Look at the letters and say each one. If the word has more than one part, say the letters part by part.
3. Write the word without looking at your book.

This method centered on intense, uninterrupted study and analysis of the word. If the pupil wrote the word correctly the first time, that was an indication he knew it. However, he was asked to write it a second and a third time to "stamp in" the correct reaction before another word was taken up.

Words from the individual lists were studied in the same manner. The teacher checked with the individual pupils to determine if the proper concepts and common meanings of the words had been established.

Reviews. Research in spelling and child development indicated that periodic review on a long list of isolated words was a waste of time. Instead, it was commonly agreed that reviews become worthwhile when there is involved an opportunity for the child to make application of the generalizations, techniques, and principles being taught. It was the assumption in this study that in a classroom organized around functional units integrated with the language arts, and the use of a textbook series based on the Rinsland study,[18] natural reviews are experienced daily.

Reference has been made in this section to the various data used in evaluation and in classroom procedures. The following section presents a more detailed treatment and interpretation of data.

Presentation and Interpretation of Data

For comparative purposes, the data in the tables for January and June are presented together.

The Standardized Tests. Data from the standardized tests indicated two important facts. First, there was a wide range of ability among the pupils; and second, there existed a general verbal and linguistic retardation of one to three grades for a group of pupils at the lower distribution level. Furthermore, the intelligence quotients did not correspond with this retardation of the pupils. This confirms the findings of Spache that vocabulary knowledge is a more significant determinant of spelling success than is intelligence, particularly in the first five grades.[19]

The California Test of Mental Maturity, Short Form.[20] This test showed the average mental age of the pupils at the beginning of the study was 145 months with a range from 115 months to 186 months. The I.Q. ranged from 76 to 136.

The Progressive Achievement Test, Form AA.[21] This test gave the average grade placement for reading vocabulary as 5.5 with a range from 2.9 to 7.9; for reading comprehension, 6.6 with a range from 4.5 to 9.5; for language, 6.5 with a range from 5.3 to 7.2.

The Gates-Russell Spelling Diagnosis Tests.[22] An analysis of the types of errors along with other data, indicated symptoms of difficulty and some definite points where instruction should be given. However, in an attempt to go deeper into the difficulties and their causes, the Gates-Russell Spelling Diagnosis Tests were given to the lower group of pupils. Results of these tests showed a general verbal and linguistic retardation of one to three years for fifteen pupils. This evidence indicated a need for a general program of word skills which would apply to all the language arts, particularly reading and spelling, and opportunities for the use of words in meaningful situations.

Misspelled Words in Grades III to VI. To determine how well the children in the school as a whole spelled, their spelling errors in the spontaneous writings were analyzed in terms of all errors made by each grade group, all the errors in the group being pooled. Tables I and II present data on this type of analysis.

Misspelled Words at Each Grade Level. In writing 35,883 running words in Grades III - VI, the children misspelled 5.95 percent of them as shown in Table I. A comparison of these results with a similar study by Swenson and Caldwell shows that these children spelled better than the children of Prairie City.[23]

Also, the increase in the number of running words and the decrease in the number of spelling errors in successive grades Indicate growth in spelling and language abilities.

Misspelled Words for the Two Sixth Grades. Table III gives an interesting comparison of the two sixth grades of the school in that the percent of total errors is much higher for the "other" sixth grade than for the one described in this study. On the basis of standardized tests and previous teacher judgment, the two groups were comparatively equal in ability at the beginning of the school term. It has already been pointed out, however, that vocabulary knowledge is a more significant determinant of spelling ability than is intelligence, particularly in the first five grades.[24] Some other causative factors could be differences in classroom curriculum, methods, attitudes, motivation, and provisions for written expression.

With the misspelling of only 3.2 words per hundred in writing 5,896 words, it appears that the experimental group had reached a rather high level of achievement in spelling by the middle of the sixth grade. By comparison, this sixth grade in misspelling 3.2 percent of the total running words spelled much better than the Prairie City sixth graders who misspelled 5.8 percent of the total running words.[25]

Misspelled Words in the Experimental Group. In order to meet the spelling needs of children, it is necessary to know which words are misspelled most frequently in the writings of children. Tables III, IV, V, VI,VII and VIII are pertinent to this.

Different Errors and Repetitions. 'The proportion of different and repeated errors are summarized in two ways in Table III. In Columns 3 and 4 of Table III, the errors of different pupils are pooled into a grade summary. When a certain misspelling of a word was tabulated for the first time, that error was included in the 69.2 percent (different words misspelled). When, later, that same word was misspelled by the same child or another child, that same error was recorded among the 30.8 percent (repetition of errors). Thus, 69.2 percent of the spelling errors in January were different errors while 30.8 percent were repetitions of errors already included once each in the 69.2 percent. Likewise for June June, 66.2 percent of the spelling errors were different for the grade while 33.8 percent were repetitions of errors already included once each in the 66.2 percent.

Data concerning the proportion of different and repeated errors are summarized in terms of the errors of individual children in Columns 5 and 6 of Table III. That is, if a certain child misspelled a particular word once, it was counted among different words misspelled, even though other children in the same grade may have misspelled the same word. An error was counted a repetition of the same error in Column 6 only when the same child made the same error.

Data in Table III indicate two important facts concerning the proportion of different and repeated errors. First, most of the errors are different individual errors; and second, the errors repeated most frequently are of two sources, namely, errors repeated by different pupils, and individual repetitions of the same errors. A disproportionate share of the errors are repetitions of

TABLE I

PROPORTION OF MISSPELLED WORDS AT EACH GRADE LEVEL IN SPONTANEOUS WRITINGS OF PUPILS IN GRADES III-IV, JANUARY, 1951

Grade	No. of Pupils	Running Words	General Words, Including Proper Nouns		Unintelligible Words		Total Errors	
			No.	Percent	No.	Percent	No.	Percent
III	74	5,269	456	8.63	25	0.47	481	9.1
IV	100	7,617	559	7.38	17	.22	576	7.6
V	84	11,825	587	4.93	20	.17	606	5.1
VI	81	11,172	450	4.03	22	.19	472	4.22
III-VI	339	35,883	2,052	5.7	84	.23	2,135	5.95

TABLE II

PROPORTION OF MISSPELLED WORDS IN SPONTANEOUS WRITINGS OF THE TWO SIXTH GRADES IN JANUARY

Grade	No. of Pupils	Running Words	General Words, Including Proper Nouns		Unintelligible Words		Total Errors	
			No.	Percent	No.	Percent	No.	Percent
VI Other	40	5,277	268	5.08	14	.26	282	5.4
VI Exp.	41	5,895	182	3.08	8	.13	190	3.2

TABLE III

PERCENTS OF DIFFERENT AND OF REPEATED ERRORS IN SPONTANEOUS WRITINGS OF PUPILS FOR JANUARY AND JUNE

Date	Total No. of General Errors	Percentage for Grade Group		Percentage for Individual Pupils	
		Different Words Misspelled	Repetition of Errors	Different Words Misspelled	Repetition of Errors
January	182	69.2	30.8	80.0	20.0
June	234	66.2	33.8	77.0	23.0

TABLE IV

126 DIFFERENT WORDS MISSPELLED AND TOTAL ERRORS FOR JANUARY IN SPONTANEOUS WRITINGS OF PUPILS

No. of Misspellings per Word	Number of Different Words Misspelled		Total Errors	
	Frequency	Cumulative Frequency	Frequency	Cumulative Frequency
1	92	92 (73%)	92	92 (51%)
2	20	112 (90%)	40	132 (73%)
3	8	120 (95%)	24	156 (86%)
4	4	124 (98%)	16	172 (95%)
5	2	126 (100%)	10	182 (100%)

TABLE V

155 DIFFERENT WORDS MISSPELLED AND TOTAL ERRORS FOR JUNE IN SPONTANEOUS WRITINGS OF PUPILS

No. of Misspellings	Number of Different Words Misspelled		Total Errors	
	Frequency	Cumulative Frequency	Frequency	Cumulative Frequency
1	111	111 (72%)	111	111 (47%)
2	26	137 (88%)	52	163 (69%)
3	10	147 (95%)	30	193 (82%)
4	3	150 (97%)	12	205 (88%)
5	2	152 (98%)	10	215 (92%)
6	2	154 (99%)	12	227 (97%)
7	1	155 (100%)	7	234 (100%)

TABLE VI

126 DIFFERENT WORDS MISSPELLED IN CHILDREN'S SPONTANEOUS WRITINGS FOR JANUARY AND OCCURRENCES IN PUPILS' OWN VOCABULARY SPELLERS, FITZGERALD LIST, AND DOLCH LIST

Word	No. of Times Misspelled	Individual Repetitions	Pupils' Own Vocabulary Spellers Grade Placement	Fitzgerald List	Dolch List
left	5	3	2	F	D
reindeer	5	4	4	-	-
excitement	4	2	6	F	-
lawn	4	3	6	F	D
their	4	0	2	F	D
when	4	2	2	F	D
Christmas	3	1	2	F	D
fright	3	2	6	F*	-
friends	3	2	3	F	D*
judges	3	2	7	F*	D*
swimming	3	2	3	F	D
there	3	1	2	F	D
too	3	0	2	F	D
where	3	2	2	F	D
announcer	2	0	3*	F*	D*
airplanes	2	0	3*	F*	D*
again	2	0	2	F	D
beaver	2	1	5	-	-
brought	2	0	2	F	D
bears	2	1	2	F	D*
couldn't	2	1	3	F	D
didn't	2	0	3	F	D
decided	2	1	2	F	D
hitch	2	1	6	-	-
haven't	2	0	5	F	D
picnic	2	1	3	F	D
same	2	0	3	F	D
so	2	1	2	F	D
Saturday	2	1	3	F	D
stay	2	1	3	F	D
since	2	0	5	F	D
Santa	2	0	2	F	D
they're	2	0	5	F	D
wait	2	0	3	F	D
asylum	1	0	-	-	-
attention	1	0	5	F	-
a lot**	1	0	2	F	D
a time**	1	0	2	F	D
a half**	1	0	3	F	D
almost	1	0	3	F	D
built	1	0	4	F	D
before	1	0	2	F	D
better	1	0	2	F	D
barking	1	0	5	F*	D*

* Indicates derived form or base word in lists.

** Indicates combined word misspelling. Most difficult one used for grade placement.

TABLE VI (Continued)

126 DIFFERENT WORDS MISSPELLED IN CHILDREN'S SPONTANEOUS WRITINGS FOR JANUARY AND OCCURRENCES IN PUPILS' OWN VOCABULARY SPELLERS, FITZGERALD LIST, AND DOLCH LIST

Word	No. of times Misspelled	Individual Repetitions	Pupils' Own Vocabulary Spellers Grade Placement	Fitzgerald List	Dolch List
bed	1	0	2	F	D
berries	1	0	4	F	D
beat	1	0	4	F	D
completely	1	0	7	F*	-
close	1	0	3	F	D
chimney	1	0	3	F	D
caught	1	0	3	F	D
coming	1	0	2	F	D
corner	1	0	4	F	D
cheese	1	0	4	F	D
did not**	1	0	3	F	D
dangerous	1	0	4	F	D
daughter	1	0	4	F	-
dodge	1	0	-	-	D
description	1	0	7	F	-
Eve	1	0	-	-	-
fined	1	9	-	-	-
find	1	0	2	F	D
finally	1	0	4	F	D
for	1	0	2	F	D
fight	1	0	3	F	D
fighting	1	0	3*	F	D*
finish	1	0	4	F	D
floated	1	0	6*	F*	D*
forgetful	1	0	3*	F*	D*
growl	1	0	-	-	D
get	1	0	2	F	D
guide	1	0	7	F	D
heard	1	0	3	F	D
hurrying	1	0	4*	F*	D*
have to**	1	0	2	F	D
haven't	1	0	5	F	D
inner tube**	1	0	6	-	-
if	1	0	2	F	D
know	1	0	2	F	D
laid	1	0	4	F	D
minutes	1	0	4	F	D*
mad	1	0	3	F	D
must	1	0	3	F	D
merry	1	0	3	F	D
money	1	0	3	F	D
Martin	1	0	-	-	-
machine	1	0	3	F	-
months	1	0	3	F	D*
noise	1	0	3	F	D
no	1	0	2	F	D
offered	1	0	6	F	D*
overboard	1	0	8	-	-

* Indicates derived form or base word in lists

** Indicates combined word misspelling. Most difficult one used for grade placement.

TABLE VI (Continued)

126 DIFFERENT WORDS MISSPELLED IN CHILDREN'S SPONTANEOUS WRITINGS FOR JANUARY AND OCCURRENCES IN PUPILS' OWN VOCABULARY SPELLERS, FITZGERALD LIST, AND DOLCH LIST

Word	No. of times Misspelled	Individual Repetitions	Pupils' Own Vocabulary Spellers Grade Placement	Fitzgerald List	Dolch List
ouch	1	0	-	-	-
prison	1	0	6	F	D
porch	1	0	3	F	D
palm	1	0	-	-	D
picked	1	0	4	F	D*
road	1	0	3	F	D
rubbed	1	0	7	-	D*
repairing	1	0	8*	-	-
Rudolph	1	0	-	-	-
reindeer	1	0	4	-	-
rhyme	1	0	-	-	-
remembered	1	0	5	F	D*
second	1	0	2	F	D
scared	1	0	4	F	D*
spilt	1	0	-	-	-
scattered	1	0	7	-	D*
smeared	1	0	-	-	-
supper	1	0	3	F	D
sleigh	1	0	3	F	D
squad	1	0	-	-	-
then	1	0	2	F	D
through	1	0	3	F	D
tomorrow	1	0	3	F	D
tried	1	0	3	F	D
they	1	0	2	F	D
think	1	0	2	F	D
turned	1	0	3	F	D*
throw	1	0	3	F	D
tale	1	0	8	-	-
the	1	0	2	F	D
teeth	1	0	3	F	D
wondered	1	0	4	F	D*
where	1	0	4	F	D
weren't	1	0	5	F	D

* Indicates derived form of base word in lists.

Total different words misspelled--126 Total errors--182 Total individual repetitions--36

Total in Vocabulary Speller, Grades I-VI, 2,726 words, 102 (81%)

Total in Vocabulary Speller, Grades I-VIII, 4,067 words, 112 (90%)

Total in Fitzgerald List of 2,650 words, 101 (80%)

Total in Dolch List of 2,000 words, 96 (76%)

TABLE VII

155 DIFFERENT WORDS MISSPELLED IN CHILDREN'S SPONTANEOUS WRITINGS FOR JUNE AND OCCURRENCES IN PUPILS' OWN VOCABULARY SPELLERS, FITZGERALD LIST, AND DOLCH LIST

Word	No. of times Misspelled	Individual Repetitions	Pupils' Own Vocabulary Spellers Grade Placement	Fitzgerald List	Dolch List
father	7	5	2	F	D
piano	6	5	3	F	D
where	6	3	2	F	D
their	5	0	2	F	D
went	5	3	2	F	D
Blue Jay**	4	3	-	-	-
hold	4	3	3	F	D
there	4	3	2	F	D
experiences	3	1	5*	F*	-
from	3	1	2	F	D
happily	3	0	7	-	-
sheriff	3	2	-	-	-
squirrel	3	2	4	F	D
tried	3	1	3	F	D
thought	3	1	3	F	D
to	3	1	2	F	D
week end**	3	2	3	F	D
whales	3	2	7*	-	-
babies	2	1	3	F	D
baseball	2	0	3	F	D
boat	2	1	2	F	D
crawled	2	1	5	-	D*
eagle's	2	1	5*	-	-
every	2	0	2	F	D
for	2	0	2	F	D
fellowship	2	1	-	-	-
front	2	1	3	F	D
finally	2	0	4	F	D
his	2	0	2	F	D
loose	2	1	4	F	-
marshmallows	2	1	6	-	-
memorial	2	1	-	-	-
neighbor	2	1	6	F	D
planned	2	0	6	F	D*
rolls	2	1	3*	F*	D*
really	2	0	4	F	-
swimmer	2	1	4*	F*	D*
stopped	2	0	3*	F	D
too	2	0	2	F	D
truly	2	1	3	F	D
thought it	2	1	4	F	D
then	2	0	2	F	D

* Indicates derived form of base word in lists.

** Indicates combined word misspelling. Most difficult one used for grade placement.

TABLE VII (Continued)

155 DIFFERENT WORDS MISSPELLED IN CHILDREN'S SPONTANEOUS WRITINGS FOR JUNE AND OCCURRENCES IN PUPILS' OWN VOCABULARY SPELLERS, FITZGERALD LIST, AND DOLCH LIST

Word	No. of times Misspelled	Individual Repetitions	Pupils' Own Vocabulary Spellers Grade Placement	Fitzgerald List	Dolch List
upon	2	1	3	F	D
which	2	1	3	F	D
altogether	1	0	6	F	-
ankles	1	0	6*	F*	D
any others**	1	0	2	F	D
ate	1	0	2	F	D
ashore	1	0	7	-	-
armies	1	0	7	-	-
a boat**	1	0	2	F	D
bowl	1	0	4	F	D
brush	1	0	4	F	D
brought	1	0	2	F	D
couldn't	1	0	3	F	D
corral	1	0	8	-	-
curiosity	1	0	7	-	-
cooked	1	0	3	F	D*
captured	1	0	5	-	D*
coming	1	0	2	F	D
dived	1	0	7*	F*	-
dumb	1	0	-	-	-
dressed	1	0	3	F	D*
didn't	1	0	3	F	D
don't	1	0	2	F	D
divide	1	0	5	F	D
diamond	1	0	7	F	-
dead	1	0	4	F	D
disappointed	1	0	6	F	-
Epson	1	0	-	-	-
friend	1	0	2	F	D
fighting	1	0	3*	F	D*
first	1	0	2	F	D
grazing	1	0	7	-	-
gnaw	1	0	-	-	-
gotten	1	0	2*	F	D*
great	1	0	3	F	D
got	1	0	2	F	D
half	1	0	3	F	D
have	1	0	2	F	D
here's	1	0	2	F	D
happened	1	0	4	F	D*
height	1	0	7	F	-
howl	1	0	5	-	-
had	1	0	2	F	D
him	1	0	2	F	D
hundred	1	0	3	F	D
haunted	1	0	7	-	-
hopped	1	0	6*	F*	D*

*Indicates derived form of base word in lists.

TABLE VII (Continued)

155 DIFFERENT WORDS MISSPELLED IN CHILDREN'S SPONTANEOUS WRITINGS FOR JUNE AND OCCURRENCES IN PUPILS' OWN VOCABULARY SPELLERS, FITZGERALD LIST, AND DOLCH LIST

Word	No. of times Misspelled	Individual Repetitions	Pupils' Own Vocabulary Spellers Grade Placement	Fitzgerald List	Dolch List
inside	1	0	3	F	D
invaded	1	0	-	-	-
jitterbug	1	0	-	-	-
knew	1	0	3	F	D
kept	1	0	3	F	D
kidnapped	1	0	8	-	-
Knox	1	0	-	-	-
knob	1	0	8	-	-
know	1	0	2	F	D
language	1	0	4	F	D
listen	1	0	4	F	D
loaves	1	0	-	-	-
morning	1	0	2	F	D
myself	1	0	4	F	-
much	1	0	2	F	D
months	1	0	3*	F*	D*
million	1	0	5	F	D
moccasins	1	0	-	-	-
minutes	1	0	4	F	D*
magic	1	0	5	-	-
nice	1	0	2	F	D
Negroes	1	0	6*	F*	D*
noise	1	0	4	F	D
no	1	0	2	F	D
owned	1	0	5	F	D*
o'clock	1	0	3	F	D
off	1	0	2	F	D
open	1	0	3	F	D
potato	1	0	4	F	D
picked	1	0	4	F	D
rabbits	1	0	2*	F	D*
ribs	1	0	8*	-	-
rattlesnakes	1	0	7*	-	-
stayed	1	0	4	F	D
scared	1	0	4	F	D*
steps	1	0	4	F	D
she would**	1	0	2	F	D
showed	1	0	3	F	D*
surprised	1	0	4	F	D*
said	1	0	2	F	D
silver	1	0	4	F	D
stunned	1	0	-	-	-
spotted	1	0	4*	F*	D*
speeded	1	0	6*	F*	D*
stomach	1	0	5	F	D
swimming	1	0	3	F	D
them	1	0	2	F	D
think	1	0	2	F	D

*Indicates derived form of base word on lists.

** Indicates combined word misspelling. Most difficult one used for grade placement.

TABLE VII (Continued)

155 DIFFERENT WORDS MISSPELLED IN CHILDREN'S SPONTANEOUS WRITINGS FOR JUNE AND OCCURRENCES IN PUPILS' OWN VOCABULARY SPELLERS, FITZGERALD LIST, AND DOLCH LIST

Word	No. of times Misspelled	Individual Repetitions	Pupils' Own Vocabulary Spellers Grade Placement	Fitzgerald List	Dolch List
times	1	0	2	F	D
they	1	0	2	F	D
taffeta	1	0	-	-	-
toilet	1	0	-	-	-
talking	1	0	5	F	D*
until	1	0	3	F	D
unoccupied	1	0	-	-	-
uncle	1	0	3	F	D
village	1	0	4	F	D
writing	1	0	3	F	D
wiggled	1	0	-	-	-
walked	1	0	3	F	D*
wheels	1	0	4*	F	D*
wiener	1	0	8	-	-
while	1	0	3	F	D
worms	1	0	4*	F*	D*
yacht	1	0	-	-	-
zoom	1	0	-	-	-

* Indicates derived form of base word in lists.
Total different words misspelled--155 Total errors--234 Total Individual repetitions--53
Total in Vocabulary Spellers, Grades I-VI, 2,726 words, 121 (78%)
Total in Vocabulary Spellers, Grades I-VIII, 4,067 words, 137 (89%)
Total in Fitzgerald List of 2,650 words, 118 (77%)
Total in Dolch List of 2,000 words, 110 (71%)

TABLE VIII

GRADE PLACEMENT, NUMBER AND PERCENT IN PUPILS' OWN VOCABULARY SPELLERS OF DIFFERENT MISSPELLED WORDS IN SPONTANEOUS WRITINGS FOR JANUARY AND JUNE

	Grade							Not in Speller	Total
	2	3	4	5	6	7	8		
January	32(25%)	35(28%)	17(13%)	10(8%)	8(6%)	6(5%)	4(3%)	14(11%)	126
June	41(26%)	33(21%)	27(17%)	11(7%)	8(6%)	11(7%)	5(3%)	18(12%)	155

TABLE IX

PROPORTION OF MISSPELLED WORDS FOR SIXTH GRADE EXPERIMENTAL GROUP IN SPONTANEOUS WRITINGS IN JANUARY AND JUNE

Date	No. of Pupils	Running Words	General Words		Unintelligible Words		Total Errors	
			No.	%	No.	%	No.	%
January	41	5895	182	3.08	8	.13	190	3.2
June	41	7764	234	3.0	0	.00	234	3.0

TABLE X

DATA ON INFORMAL SPELLING INVENTORY SHOWING LEVELS OF ACHIEVEMENT AND MEANS FOR JANUARY AND JUNE

Date	Form	Number of Pupils at Different Grade Levels of Achievement. Grade:								Median Level	Mean Level
		1	2	3	4	5	6	7	8		
1-1-51	M	1	4	11	3	10	8	3	1	5	4.4
6-12-51	A	0	0	4	3	12	10	7	5	6	5.7

errors by individual children.

An interesting fact concerning the data for January and June is that the percents of different words misspelled for the grade group and for individual pupils are less in June than in January although 1,869 more words were written.

Frequency of Occurrence of Misspelled Words. Data in Tables IV and V offer specific information on the source of spelling errors. The first line of Table IV is read as follows: There were 92 different words which were misspelled only once in all the stories. These accounted for 73 percent of the 126 different words misspelled, but they comprised only 51 percent of the 182 total spelling errors. It is obvious that the few words represented at the bottom of Tables IV and V account for a disproportionately high percentage of the errors in the children's writing.

Tables VI and VII indicate the specific sources of these errors as repetitions of errors by different children, repetitions of errors by the same child, or a combination of both.

Different Words Misspelled. Different words misspelled, the number of times misspelled, individual repetitions, and occurrences in Pupils' Own Vocabulary Spelling Series,[26] A Basic Life Spelling Vocabulary,[27] and Dolch's 2,000 Commonest Words for Spelling,[28] are shown in Tables VI and VII, which are self-explanatory. How well each list met the needs of the pupils is given at the bottom of the tables in percents. If derived forms had been omitted, the percentages would have been somewhat less.

A further analysis of the total misspellings of these different words shows that in January 85% of the total errors occurred in words in the Pupil's Own Vocabulary Spellers, Grades I-VI, consisting of 2,726 words, and 92% occurred in words in the spelling series, Grades I-VIII of 4,067 words; that 82% occurred in words in the Fitzgerald List of 2,650 words; and that 79% occurred in words in the Dolch List of 2,000 Commonest Words.

Likewise for June, 80% of the total errors occurred in the Pupils' Own Vocabulary Spellers, Grades I-VI and 89% in the spelling series; that 78% occurred in the Fitzgerald List; and that 73% occurred in the Dolch List. These tables indicate that with the exception of unusual words, proper nouns, derived forms, and words of low frequency of usage, the writing needs of the pupils are adequately met for the most part in the lists.

These tables show that the words misspelled most frequently were either repetitions of errors by different children on "demon words", or repetitions of errors by the same child on a particular word.

Occurrence of Misspelled Words in Pupils' Own Vocabulary Spellers for January and June. An examination of Table VIII reveals two important facts, namely, that the spelling needs of the pupils cannot be met by a single spelling textbook, and that it cannot be assumed that the spelling words of the previous grades have been mastered by the pupils. For example, 74 percent of the different words misspelled in January are in a grade placement below the sixth grade, and 71 percent of the different words misspelled in June are in a grade placement below the sixth grade. Furthermore, 53 percent and 47 percent of the different misspelled words in January and June respectively, are placed in the second and third grades.

Proportion of Misspelled Words in January and June. Evidence on the effectiveness of the pupils' writings in January and June is given in Table IX. The reduction in pupil errors of .2 percent over the period of five months seems insignificant in itself. However, with the increased use of 1869 words and the probability of error among writings of sixth grade pupils with as wide a range of ability as this group had, it is doubtful if the percentage of errors could have been reduced much more in a five months period. By comparison again, the Prairie City sixth grade pupils made 2.8 percent more errors than did this sixth grade group. Also, this group would be equivalent to half-way between the seventh and eighth grade of the Prairie City school children.[29]

Informal Spelling Inventory Analysis. One of the best measurements of achievement for a particular group is the informal inventory because it evaluates the pupils on materials which they have had an opportunity to learn.

Temple University Informal Spelling Inventory.[30] Table X is self-explanatory. During the five months of the experiment, the pupils made a mean gain of 1.3 grades. The mean gain was actually higher because five pupils achieved beyond the eighth grade.

Teacher-Prepared Informal Inventory. The results of the teacher-prepared inventory as given in Table XI are similar to that of the one just described. It is not possible to give the mean because tests were not given beyond the sixth grade. Eight pupils scored beyond the sixth grade in January and fourteen pupils scored beyond the sixth grade in June. It is interesting to note, also, that in January fifteen pupils had an instructional level of second and third grade. In June this was reduced to four in number indicating much progress in the group which caused a majority of errors in the pupils' writings.

Summary and Conclusions

The purpose of this study was (a) to determine the spelling needs of the pupils of a sixth grade class through classroom experimentation and research, and (b) to suggest procedures for spelling instruction to meet those needs by utilizing the implications of scientific investigation and

TABLE XI

DATA ON TEACHER PREPARED INFORMAL SPELLING INVENTORY SHOWING LEVELS OF ACHIEVEMENT FOR JANUARY AND JUNE

Date	Number of Pupils at Different Grade Levels						
	1	2	3	4	5	6	Beyond
1-1-51	0	5	10	5	8	5	8
6-12-51	0	1	3	3	10	10	14

This presentation and interpretation of data is given for a particular school and a particular grade of that school. The evidence of this data may or may not be applicable to other classroom situations.

expert opinion.

To determine the spelling needs of those pupils, four main sources of data were used: 1) standardized tests, 2) informal inventories in spelling and reading, 3) analysis of spelling errors in pupils' spontaneous writings, and 4) the research and theory in spelling and related areas.

Initial analysis of the above data indicated that the spelling abilities of the children varied widely and that a core of approximately 2500 words made up about 95 percent of the words which the children used in writing. Moreover, there were three groups of pupils within the total group that had similar levels of achievement and needs. To meet the needs of the pupils within these groups, spelling instruction was integrated with language arts. To provide opportunities for using spelling in meaningful situations, instruction in the content areas was organized around the functional unit or experience approach.

Among the detailed findings of this study, the following points seem to be most significant:

1. Spelling instruction, to be most effective, must be differentiated in terms of the needs of the group and of the individuals within the group.

2. Vocabulary knowledge is a more significant determinant of spelling success than is intelligence, particularly in the elementary grades.

3. A general verbal and linguistic retardation is a major cause of poor spelling.

4. Most of the spelling errors which sixth grade children make in spontaneous writing are individual errors.

5. The spelling errors repeated most frequently are either errors repeated by different pupils, or individual repetitions by the same individual pupil.

6. The words needed for sixth grade children's spontaneous writing can be met, for the most part, by a reliable core list supplemented by individual word lists checked for frequency of occurrence with a valid list.

7. The use of a single textbook for all the children in a class does not meet the needs of the pupils.

8. It cannot be assumed that mastery of spelling words and word attack skills has been achieved and maintained in previous grades.

9. The informal inventories in spelling and reading are effective aids in grouping of pupils on levels of ability.

10. Spelling ability is one of a constellation of language arts abilities that tend to reinforce each other.

11. The study-test plan of instruction is an effective procedure in instructing retarded or slow-learning pupils, and the test-study plan is an effective procedure for the accelerated pupils.

12. An analysis of children's spontaneous writings is an effective means for determining pupils' spelling needs.

13. Many pupils of the sixth grade have not developed an effective method of studying a word.

14. Incidental learning should be supplemented by direct systematic teaching, especially in the case of difficult words and words of high frequency of occurrence.

15. Children, working on levels of achievement and using spelling in meaningful situations, tend to have favorable attitudes toward spelling.

16. Adequate research, for all practical purposes, appears to be available for improved spelling instruction.

The final conclusion of the writer is that while the functional method was not necessarily better for all pupils, it was better than the conventional method for many pupils in the class. The functional method of teaching spelling in a curriculum organized as described in this study was reasonably effective in meeting the needs of the pupils. Individual differences provide opportunities for rich living. In a democratic society, these differences are regarded as assets.

FOOTNOTES

1. James A. Fitzgerald, The Teaching of Spelling (Milwaukee: The Bruce Publishing Company, 1951), p. 2.
2. California Test of Mental Maturity (Los Angeles: California Test Bureau, 1950).
3. Progressive Achievement Tests--Elementary Battery (Los Angeles: California Test Bureau, 1950).
4. Arthur I. Gates, Henry D. Rinsland, Ina C. Sartorius, and Celeste C. Pearson, The Pupils' Own Vocabulary Speller, revised edition (New York: The Macmillan Company, 1950).
5. James A. Fitzgerald, A Basic Life Spelling Vocabulary (Milwaukee: The Bruce Publishing Company, 1951).
6. Edward W. Dolch, Better Spelling (Champaign, Ill.: The Garrard Press, 1942), pp. 257-270.
7. Henry D. Rinsland, A Basic Vocabulary of Elementary School Children (New York: Macmillan Company, 1945), 636 pp.
8. William S. Gray, Basic Readers: Curriculum Foundation Series (New York, Scott Foresman Company, 1950).
9. Donald Durrell, Helen Sullivan and Helen Murphy, Building Word Power (New York: World Book Company, 1945).
10. Thorndike Century Beginning Dictionary (New York: Scott Foresman Company, 1945).
11. Temple University Informal Spelling Inventory (Reading Analysis of the Reading Clinic, Temple University, Philadelphia.
12. Gates, et al., op. cit.
13. Gates-Russell Diagnostic Spelling Tests (New York: Bureau of Publications, Teach-

ers College, Columbia University, 1937).

14. California Test of Mental Maturity, loc. cit.
15. Progressive Achievement Tests, loc. cit.
16. A. Sterl Artley, "Principles Applying to the Improvement of Spelling Ability," Elementary School Journal, 49: 137, November, 1948, pp. 137-148.
17. David H. Russell, Characteristics of Good and Poor Spellers; A Diagnostic Study, Teachers College Contribution to Education, No. 272. (New York: Bureau of Publications, Teachers College, Columbia University, 1937).
18. Rinsland, loc. cit.
19. George Spache, "Spelling Disability," Journal of Educational Research, 34: 561-586, April, 1941.
20. Loc. cit.
21. Loc. cit.
22. Loc. cit.
23. Ester J. Swenson and Charles G. Caldwell, "Spelling in Children's Letters," Elementary School Journal, 49: 224-235, December, 1948.
24. Spache, loc. cit.
25. Swenson and Caldwell, op. cit.
26. Gates, et al., loc. cit.
27. Fitzgerald, loc. cit.
28. Dolch, loc. cit.
29. Swenson and Caldwell, op. cit.
30. Loc. cit.

BIBLIOGRAPHY

Betts, Emmett A., Foundations of Reading Instruction (New York: American Book Company, 1946).

Betts, Emmett A., "Inter-relationship of Reading and Spelling," Elementary English Review, 22:13, January, 1945.

Brittain, Francis J., and Fitzgerald, James A., "The Vocabulary and Spelling Errors of Second Grade Children's Themes," Elementary English Review, 19:43-50, February, 1942.

California Test of Mental Maturity, (Los Angeles: California Test Bureau).

Dolch, Edward W., Better Spelling (Champaign, Illinois: The Garrard Press, 1942), 270 pp.

Durrell, Donald D.; Sullivan, Helen Blair, and Murphy, Helen A., Building Word Power, (New York: World Book Company, 1945), 102 pp.

Durrell, Donald D., Improvement of Basic Reading Abilities, (New York: World Book Company, 1940), 407 pp.

Fitzgerald, James A., A Basic Life Spelling Vocabulary, (Milwaukee: The Bruce Publishing Company, 1951), 161 pp.

Fitzgerald, James A., The Teaching of Spelling, (Milwaukee: The Bruce Publishing Company, 1951) 233 pp.

Fox, Williams H., and Eaton, Merrill T., Analysis of Spelling Proficiency of 82,833 Pupils in Grades 2 to 8 in 3,547 Teaching Units of the City Schools of Indiana, Bulletin of the School of Education, Indiana University, Vol. XXII, No. 2 (Bloomington: Bureau of Cooperative Research and Field Service, Indiana University, 1946).

Fox, William H., Spelling Proficiency in Township Schools in Indiana, Bulletin of the School of Education, Indiana University, Vol. 23, No. 1, (Bloomington: Division of Research and Field Services, Indiana University, 1947).

Gates, Arthur I.; Rinsland, Henry D.; Sartorious, Ina C.; and Peardon, Celeste C., The Pupils' Own Vocabulary Speller, revised edition, (New York: The Macmillan Company, 1950).

Gates, Arthur I., Spelling Difficulties in 3,876 Words, (New York: Bureau of Publications, Teachers College, Columbia University, 1937), 166 pp.

Gates-Russell Diagnostic Spelling Tests(New York: Bureau of Publications, Teachers College, Columbia University, 1937).

Gray, William S., Basic Readers: Curriculum Foundation Series, (New York: Scott, Foresman Company, 1950).

Gray, William S., Developing Word Attack Skills, (New York: Scott, Foresman Company, 1947.

Gray, William S., On Their Own in Reading, (New York: Scott, Foresman Company, 1948) 268 pp.

Gray, William S., et al., The Guidebook for Basic Readers: Curriculum Foundation Series, (New York: Scott, Foresman Company, 1950).

Hildreth, Gertrude, "New Light on the Spelling Problem," Elementary English, 25:201-207, April, 1948.

Horn, Ernest, "Spelling," Encyclopedia of Educational Research, revised edition, 1950, pp. 1247-1264.

Lee, J. Murray, and Lee, Doris May, The Child and His Curriculum, (New York: Appleton-Century-Crofts Inc., 1950), 710 pp.

McKee, Paul, Language in the Elementary School, revised edition, (Boston: Houghton Mifflin Company, 1939), pp. 329-440.

Progressive Achievement Tests-Elementary Battery, (Los Angeles: California Test Bureau, 1950).

Rinsland, Henry D., A Basic Vocabulary of Elementary School Children, (New York: Macmillan Co., 1945), 636 pp.

Russell, David H., Characteristics of Good and Poor Spellers, Teachers College Contribution to Education, No. 727 (New York: Bureau of Publications, Teachers College, Columbia University, 1937), 103 pp.

Spache, George, "Spelling Disability," Journal of Educational Research, 34:561-586, April, 1941.

Sterl, Artley A., "Principles Applying to the Improvement of Spelling Ability," Elementary School Journal, 49:137-148, November, 1948.

Swenson, Estes J., and Caldwell, Charles G., "Spelling in Children's Letters," Elementary School Journal, 49:224-235, December, 1948.

Wilson, Louis Ada, "Children's Spelling Needs and Their Implications for Classroom Procedures," Elementary School Journal, 47: 98-102, October, 1946.

1.8.70
J304

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 1, September 1960)

PUBLIC ATTITUDES TOWARD EDUCATION

DONALD B. BLACK et al.*
University of Alberta

THE SEVERE criticism to which our educational system has been subjected recently has resulted directly in the establishment of a number of public investigating groups to examine educational matters. In Canada, these groups are established as Royal Commissions in the particular provinces concerned. At present, three are sitting or have just concluded sittings in four of the western provinces. While many educators view such investigations as indicative of renewed public interest in their schools, it is the purpose of this study to attempt to establish the degree of this interest by determining the extent to which a small segment of the Alberta population is aware of the existence of the Alberta Royal Commission on Education and some of this sample's attitudes towards the provincial educational system in general.

Method

On August 5, 1958, between 7 and 9 p.m., forty-four interviewers telephoned 831 private homes in the Edmonton area. The numbers called were chosen at random from the numbers listed in the back of the Edmonton Telephone Directory and were checked in the directory proper to ensure that each number was for a residence telephone. Of the 831 calls made, 225 residents did not answer or their lines were busy. These numbers were called a second time after a half-hour interval. Forty-four of those answering were not eligible for inclusion in the survey because they were under the age of eighteen, resident outside the province, or unable to speak English. Fourteen interviews were not completed because of such interruptions as the respondents answering the door, feeding the baby, or visitors arriving. These cases were excluded from the sample. Sixty-nine respondents refused to be interviewed, some because they thought the call was a device to sell magazines, encyclopedias, insurance, correspondence courses, vacuum cleaners, or to solicit religious contributions. Others said that they could not be bothered. It should be noted here, that the use of the psuedo-telephone survey as a sales promotional device constitutes a serious threat to the validity of studies made by legitimate research groups. Even among the interviews completed, the prevalence of the ''What's your gimmick?'' attitude at the end of the interview was marked. On the whole, however, the interviewers reported excellent cooperation from the 158 men and 321 women who answered the interviewer's questions. In every instance, the anonymity of the respondent was protected.

The calls took the form of a structured interview in which answers to specific questions were sought. The interview had three sections: data about the respondent, such as age, formal education, children attending school, presence of newspapers, radio, and television in the home; questions about the Alberta Royal Commission on Education; and last, questions about the Provincial educational program in general. The responses were recorded on a specially prepared data sheet which also served as the interview outline. Notwithstanding the presence of the form before them, some of the interviewers on occasion neglected to ask some of the questions and this was duly recorded.

Analysis of Data

The recorded replies on interview forms were classified and coded for punching into IBM cards. Subsequent analysis of the data was done on the IBM type 101 Electronic Statistical Machine.

It was a hypothesis of this study that there are three factors that would influence responses to the various questions asked. The first of these is called ''parental status.'' The 479 respondents were divided into five categories based on information given in the interview. These five categories are: 1) single, 2) married with no children (N. C), 3) married but with pre-school age children only (P. S. O.), 4) married with children attending school in grades one through university or undergoing professional training, such as, nursing (C. A. S.), and 5) married respondents whose children had left school (C. G. U.).

---

*This study reports a project conducted by the students in the Methods of Educational Research class during the 1958 Summer School at the University of Alberta.

The second broad categorization dealt with the amount of formal education the respondent had received. Six categories were employed: 1) elementary, grades 1-8 (ELEM); 2) high school, grades 9-12 (H. S.); 3) university and/or professional training beyond high school, such as nursing (UNIV.); 4) technical or commercial training beyond high school (TECH.); 5) a group consisting primarily of European-born respondents whose equivalent educational level was not known (OTHERS), and finally, 6) a group who did not state or were not asked the level of their formal schooling by the interviewer (NO. ST.).

The third area of interest was the age of the respondents. The interviewer in the conversation asked the respondent into which of five age groups his or her age fell. For six who would not divulge their age, a not-stated category (Not Stated) was added.

The first analysis sought to establish the access of the respondent to educational information and specifically, information about the Alberta Royal Commission on Education. This dealt with such items as regular receiving of a newspaper, radio and television in the home, and membership in Home and School Associations.

Of the 479 respondents, 455 received the one local daily newspaper regularly while only eighteen received no paper. The remainder received an out-of-town English language paper only or a foreign language paper only. Four hundred and eighteen listened predominantly to one of the five local English-speaking radio stations while only four listened predominantly to the one local French station. Only six of the respondents indicated they had no radio, whereas forty-five did not state a preference or listened predominantly to foreign stations via short wave. Responses for six persons were not recorded. Concerning television, 382 had access to sets; ninety-two did not. Four responses were not noted.

The relationship between Home and School Association membership and parental status is reported in Table I. This table would suggest that in terms of membership in the Home and School Associations, slightly less than half the parents (47%) with at least one child attending school are members, whereas eleven percent of this group had belonged at one time.

It must be noted at this point that the press and radio coverage of the Alberta Royal Commission on Education has been acknowledged openly by the Commission members to be both extensive and impartial in their presentation of the briefs placed before this Commission.

It can be concluded from the above that the vast majority of the sampled population have had ready access in some form to information about the Commission and its activities and other educational matters inasmuch as this information is given by the local press and radio-television media. On the other hand, information about the provincial and local educational policy of a detailed and specific nature that one would expect to be normally made available through organizations interested directly in education and represented in this study by membership in a home and school association, has been made available to only one-third of the sampled population.

The first group of questions asked by the interviewers dealt with the Alberta Royal Commission on Education (the Cameron Commission). The opening question sought to establish awareness of the existence of the Commission. If respondents did not recognize it at first, the interviewer gave certain additional information about the Commission and any delayed recognition was recorded. These data are reported in Table II. Approximately half the respondents in the sample indicated that they had heard of the Cameron Commission but almost an equal number had not. Twenty-six recognized it after additional information was given. Clearly, the group with a university or professional formal education were, as a single group, most aware of the existence of the Commission, whereas the respondents with an elementary education were least aware of it. For the remainder of the categories, the proportion of "awares" to "not awares" was about half.

It was appreciated, however, that the mere recognition of the Royal Commission on Education was not sufficient to indicate knowledge of the Commission's activities. Those who had heard of the Commission were also asked, "What is this Commission trying to do?" The responses to this question classified by degree of correctness of answers are reported in Table III.

It is evident from this table that although 230 recognized the Cameron Commission, only slightly more than half of this group had any knowledge of the objectives of the Commission. Again, it is evident that no marked differences exist between the portions of responses of the sampled group categorized by parental status and age, but that the formal education of the respondent does bear an important relationship to knowledge of the Commission. It would appear that the degree of knowledge about the Commission varies directly with the amount of formal education. Of the group with a university or professional education, only a quarter admitting recognition of the Commission had no knowledge of it. For the remainder of the categories, this proportion was about half, with the exception of the group with only an elementary formal education. Three-quarters of this particular group who claimed knowledge of the Commission did not know what it was trying to do.

The respondents who recognized the Royal Commission either on presentation of its title or after receiving partial information about it, were asked what they remembered hearing about the

TABLE I

RELATIONSHIP OF MEMBERSHIP IN A HOME AND SCHOOL ASSOCIATION TO PARENTAL STATUS (N =479)

	Membership in a Home and School Association				
Parental Status	Never Belonged	Currently Belong	Had Belonged	Response Not Indicated	Total
Single	48	--	1	29	78
Married:					
No Children	53	6	7	8	74
Children grown up	50	1	6	9	66
Preschool children only	50	2	4	10	66
Children attending school	76	91	21	7	195
Total	277	100	39	63	479

TABLE II

RELATIONSHIPS OF AGE, PARENTAL STATUS, AND EDUCATION TO STATED AWARENESS OF THE EXISTENCE OF THE ALBERTA ROYAL COMMISSION ON EDUCATION (N = 479)

Have you heard of the Alberta Royal Commission on Education or the Cameron Commission as it is often called?	Parental Status					Formal Education						Age of Respondent					
		Married															
Response Category	Single	N. C.	C. G. U.	P. S. O.	C. A. S.	ELEM.	H. S.	UNIV.	TECH.	OTHER	NO. ST.	Under 21	21-34	35-50	Over 50	Not Stated	Total
Yes	31	34	35	31	99	27	139	45	16	2	1	13	81	98	36	2	230
No	37	36	28	33	86	88	107	7	9	5	4	14	88	77	37	4	220
Yes, after clues given	9	3	3	2	9	4	11	8	2	1	--	3	7	10	6	--	26
Response Not Recorded	1	1	--	--	1	--	2	--	--	--	1	1	1	1	--	--	3
Total	78	74	66	66	195	119	259	60	27	8	6	31	177	186	79	6	479

TABLE III

RELATIONSHIPS OF AGE, PARENTAL STATUS, AND EDUCATION TO KNOWLEDGE OF ALBERTA ROYAL COMMISSION ON EDUCATION OF THOSE EXPRESSING INITIAL AWARENESS OF COMMISSION'S EXISTENCE (N = 230)

What is this Commission trying to do?		Parental Status				Formal Education						Age of Respondent					
		Married															
Response Category	Single	N. C.	C. G. U.	P. S. O.	C. A. S.	ELEM.	H. S.	UNIV.	TECH.	OTHER	NO. ST.	Under 21	21-34	35-50	Over 50	Not Stated	Total
Meaningless replies	3	3	2	2	8	4	12	1	1	--	--	1	5	9	2	1	18
No Opinion (I don't know)	11	10	14	7	24	12	44	4	4	1	1	7	25	21	13	--	66
Confused or Mistaken Identity	--	2	--	3	5	2	7	--	1	--	--	--	4	4	2	--	10
Some Knowledge (Confused Concept of Authority)	10	14	9	9	35	7	43	20	6	1	--	2	28	37	9	1	77
Thorough Knowledge	3	4	7	8	23	--	24	19	2	--	--	2	13	22	8	--	45
Response Not Indicated	4	1	3	2	4	2	9	1	2	--	--	1	6	5	2	--	14
Total	31	34	35	31	99	27	139	45	16	2	1	13	81	98	36	2	230

Cameron Commission. Categorization was done in terms of the content of their replies in relation to education and the Commission itself. Negative remarks about the Commission included such comments as "a bunch of eggheads wasting the taxpayer's money." Positive statements were represented by remarks such as, "this is a good thing and long over due." Positive and negative remarks about education materialized in such comments as "seeing what a good job the schools are doing," to "finding out how poor our schools really are." Comments relating to Commission personnel referred to specific members of the Commission or events involving specific members of the Commission. Incorrect information here, as in the previous table, dealt with those who mistook the identity of the Commission concerned, confusing it with such groups as the Liquor Commission. These data are presented in Table IV.

This table confirms the findings of Table II in that half of the respondents could not recall hearing anything about the Cameron Commission. Most frequent responses indicated a favorable attitude to the Commission and a negative attitude to education. Positive statements are not far behind in frequency. In Table IV, no clear relationship between categorization of respondents and response is evident other than that reported earlier concerning amount of formal education.

The last question asked, in regard to the Royal Commission, was "What is your frank opinion of its possible influence on Alberta education?" This was an attempt to identify the extent of the confidence in the Commission's capacity for diagnosis of, and prescription for the province's educational ills, as well as the government's acceptance of the Commission's findings. Again, the responses were categorized according to the nature of the response. With one exception, the categories are self-descriptive. The Qualified Optimism consists of those commenting that "the Commission will do a lot of good if the findings are followed up." These data are reported in Table V.

Table V shows that those expressing optimism about the value of the Royal Commission on Education outnumber those who have no opinion in this regard and far outweigh the number of those expressing pessimism. Again, respondents with university or professional education give a much higher proportion of positive opinions which are predominantly optimistic. By contrast, the bulk of the group with only elementary schooling have no stated opinions. Only one other group, married respondents with children of school age, share the same willingness to commit themselves to an opinion. On the other hand, the single group in the Parental Status category and the "under 21" group of the "Age" analysis, who obviously represent a marked overlap in categorization, showed an unwillingness to make any commitment regarding the possible value (or lack of it) of the Royal Commission on Education.

In summary, the effect of widespread reporting of the Royal Commission on Education activities has not resulted, for the sample surveyed at least, in any marked awareness of the Commission and its activities. Perhaps this reflects a general lack of interest in schools by the general public. The only factor that does seem related to interest in the Commission is the amount of formal education received by the respondent. University and professionally trained respondents are far more cognizant of the Commission's existence, aims, and objectives, and are more optimistic of its potential value than are any other group. On the other hand, the group with a formal schooling terminated at the elementary level showed a general unawareness of and lack of concern for the activities of the Commission. Surprisingly, the categorization by "parental status" of married respondents with and without children enrolled or about to be enrolled in school produced, in most instances, few differences of the magnitude reported for amount of formal education categorization of respondents. The same results were noted for the categorization by age of respondent.

The second part of the study sought to establish the attitudes of the sample surveyed towards the Provincial educational program in general. The respondents were asked first to give one feature of the educational system they thought was weakest; next, what they thought was good about the educational system; and finally, they were asked to make one suggestion that would in their opinion improve the caliber of our schools regardless of how good or bad they rated the schools.

Obviously, there was a wide variety of replies to such leading questions. Consequently, in the development of categories, it must be emphasized that these may include both positive and negative aspects, particularly in answer to the question on improving the schools. For example, in the homework category, some respondents thought that there should be more homework whereas others thought there was already too much and this should be reduced. This study attempted only to identify areas of response of the sample to the questions, not to measure the magnitude or direction of these responses.

In reply to the question asking what the respondent thought weakest about our schools, twelve categories resulted, including one for respondents who did not reply or to whom the question was not asked by the interviewer. These data are presented in Table VI.

It can be seen from this table that about a quarter of the sample had no opinions about what they felt wrong with the schools. Twenty-four, or five percent of the respondents stated they had no criticism of the schools. Dissatisfaction with the

TABLE IV

RELATIONSHIP OF AGE, PARENTAL STATUS, AND EDUCATION TO RECALL OF INFORMATION ABOUT ALBERTA ROYAL COMMISSION ON EDUCATION OF RESPONDENTS INDICATING AWARENESS OF ITS EXISTENCE (N = 256)

Have you heard of the Alberta Royal Commission on Education or the Cameron Commission as it is sometimes called?	Parental Status					Formal Education						Age of Respondents					
		Married															
Response Category	Single	N. C.	C. G. U.	P. S. O.	C. A. S.	ELEM.	H. S.	UNIV.	TECH.	OTHER	No. ST.	Under 21	21-34	35-50	Over 50	Not Stated	Total
Nothing	15	5	11	7	24	10	44	3	4	--	1	8	21	24	9	--	62
Can't recall just now	10	13	6	5	25	13	33	9	2	2	--	3	19	27	9	1	59
Negative to Commission	--	3	--	--	4	--	5	1	1	--	--	--	3	3	1	--	7
Positive to Commission	3	3	7	5	19	2	16	15	4	--	--	--	11	18	8	--	37
Negative to Education	4	5	3	2	14	--	17	10	1	--	--	3	8	12	4	1	28
Positive to Education	2	3	2	6	10	2	12	8	1	--	--	1	11	9	2	--	23
Commission Personnel	1	--	2	2	2	1	4	1	1	--	--	1	1	3	2	--	7
Incorrect Information	4	1	2	3	3	--	6	3	4	--	--	--	7	5	1	--	13
Response not indicated	1	4	5	3	7	3	13	3	--	1	--	--	7	7	6	--	20
Total	40	37	38	33	108	31	150	53	18	3	1	16	88	108	42	2	256

TABLE V

RELATIONSHIP OF PARENTAL STATUS, EDUCATION, AND AGE TO CONFIDENCE IN VALUE OF ALBERTA ROYAL COMMISSION OF RESPONDENTS INDICATING AWARENESS OF IT'S EXISTENCE (N =256)

What is your frank opinion of it's (Royal Commission's) possible influence on Alberta Education?	Parental Status					Formal Education						Age of Respondents					
		Married															
Response Category	Single	N. C.	C. G. U.	P. S. O.	C. A. S.	ELEM.	H. S.	UNIV.	TECH.	OTHER	NO. ST.	Under 21	21-34	35-50	Over 50	Not Stated	Total
Don't know	18	13	15	3	31	14	52	6	8	2	--	9	29	28	16	--	82
Extremely Optimistic	3	3	2	3	6	1	10	6	--	--	--	1	5	8	3	--	17
Optimistic	7	11	10	15	43	7	48	25	6	--	--	3	28	40	14	1	86
Pessimistic	2	6	4	1	8	1	11	7	1	1	--	--	6	12	2	1	21
Emphatically Pessimistic	3	--	3	1	3	1	6	2	1	--	--	2	2	3	3	--	10
Qualified Optimism (Proviso)	5	2	1	4	10	3	11	6	2	--	--	--	10	10	2	--	22
Neutral (Won't do any harm)	--	--	--	--	4	--	3	1	--	--	--	--	2	2	--	--	4
Response not indicated	2	2	3	4	3	4	9	--	--	--	1	1	6	5	2	--	14
Total	40	37	38	33	108	31	150	53	18	3	1	16	88	108	42	2	258

TABLE VI

RESPONSES CATEGORIZED BY PARENTAL STATUS, FORMAL EDUCATION, AND AGE OF RESPONDENTS WHEN ASKED TO INDICATE ONE WEAK FEATURE OF ALBERTA'S EDUCATIONAL PROGRAM (N = 479)

If you had to state one feature of our schools that is weakest, what would this be?	Parental Status					Formal Education						Age of Respondents					
		Married															
Response Category	Single	N. C.	C. G. U.	P. S. O.	C. A. S.	ELEM.	H. S.	UNIV.	TECH.	OTHER	NO. ST.	Under 21	21-34	35-50	Over 50	Not Stated	Total
Don't know, or couldn't say	24	18	16	25	39	43	64	5	5	2	3	8	48	41	22	3	122
No criticism of schools	6	4	4	2	8	9	13	1	--	--	1	2	12	5	4	1	24
Dissatisfaction over curriculum	14	12	11	11	37	10	49	15	10	1	--	6	31	36	12	--	85
Dissatisfaction over teaching methods	6	13	4	4	27	7	34	10	3	--	--	4	18	24	7	1	54
Division of home and school responsibilities	1	3	1	1	1	1	4	1	--	--	1	--	4	1	2	--	7
Extra-curricular activities	2	2	3	1	6	6	4	2	2	--	--	--	5	7	2	--	14
Standard of achievement	9	4	8	6	13	9	18	8	2	3	--	1	15	13	11	--	40
Standard of teacher qualifications	2	3	2	3	10	2	15	3	--	--	--	3	9	7	1	--	20
Crowded conditions	--	--	--	1	6	2	4	1	--	--	--	--	2	5	--	--	7
Lack of discipline	9	11	11	7	35	22	39	8	3	--	--	4	21	32	14	1	72
Miscellaneous replies	5	4	5	5	10	5	15	5	2	2	--	3	10	12	4	--	29
Response not indicated	1	--	1	--	3	3	--	1	--	--	1	--	2	3	--	--	5
Total	78	74	66	66	195	119	259	60	27	8	6	31	177	186	79	6	479

present curriculum was the most common complaint. Lack of discipline in the schools ranked second. Standards of achievement seemed more important to the respondents since standards received twice as many responses as did concern for teacher qualifications. However, less than ten and five percent of the sample indicated standards of achievement or teacher qualifications, respectively, as primary concerns. By contrast, the division of home-school responsibilities which included religious training and complaints about crowded conditions were noted only by seven respondents in each case from the total sample of 479.

The major relationship evident in Table VI lies between the amount of formal education and response. It is clearly evident here that the amount of education does relate directly to the willingness or ability to respond. It is to be noted that respondents with an elementary education place far more stress on the lack of discipline than do those with university or professional training, while this latter group shows greater concern for the curriculum and teaching methods.

The responses to the item which asked the respondents to state one feature they thought was good about the schools is presented in Table VII.

The results of the "one good feature" item indicated that almost three-eighths of the sample could not state one feature they thought was good about the schools. Moreover, an additional eighteen felt they could say nothing good about Alberta's educational program, whereas eighty-seven or over eighteen percent of the respondents felt the schools were doing an "all-round good job." Apart from such general responses, forty-six persons found occasion to make comments about teachers and teaching methods as good features of the present schools. On the other hand, school extra-curricular activities and the schools' provision for individual differences found mention by the least number of respondents. As noted earlier, the nature of the response bears closer relationship to the amount of formal education of the respondent than either parental status or age of the respondent.

Mark Twain once remarked that "Everybody talks about the weather but nobody does anything about it." The last question gave respondents opportunity to make one suggestion that would in their opinion improve the present educational system. It was here more than in either of the previous questions that the categorization used is meaningless in regard to the positive or negative aspect of the response. However, when the number of responses warranted, an attempt was made to differentiate between positive and negative responses within a single category. In all, thirteen categories resulted, including one for responses not reported. While the majority are self-descriptive, two need further clarification. The "Administrative Arrangements" category includes scholarships, financial equalization, and uniformity of curriculum from province to province and from school to school. The category on Moral Attitudes and Values includes responses that indicate the need for religious training, the desirability of fighting communism, the importance of teaching respect for elders, and like responses. It should be noted that unlike the previous two sets of answers, all responses were able to "fit" into the created categories without recourse to a miscellaneous category. These data are presented in Table VIII.

The number of respondents who didn't know, couldn't say, or were so satisfied with the schools that they would suggest nothing, accounted for over a third of the sampled population. Improvement of teacher qualifications and teaching methods was cited most frequently as the one way to improve our schools, with teaching Moral Attitudes and Values least frequently mentioned. As in previous analyses in this study, the amount of formal education seems to bear the greatest relationship to response than either parental status and age.

Limitations of the Study

No attempt has been made to analyse the reported data exhaustively for relationships other than those that are readily evident. The responses of the interviewees have been assumed to be valid expressions of their opinions and have been correctly recorded by the interviewers. The sampling technique obviously limits the surveyed population to those who possess telephones with listed numbers, and significantly more women than men answered the telephone at the time of night the calls were placed. Concern is expressed for dwellers in newly developed areas of the city which are presently served poorly by telephones. Many of these areas are populated predominantly by young families with pre-school and school-age children. It is believed that socio-economic status bears little direct relationship to possession of a telephone except for the extreme low end of the scale. Moreover, if we assume that the majority of respondents have lived in the Edmonton area or even in Alberta for any length of time, their opinions about Alberta education will be based predominantly on the public, separate and private schools of the Edmonton area. Obviously, the same survey made in an area other than the one used in this study might produce a different configuration of results. The readily-documented superior qualifications of city teachers as a group compared with those of teachers in rural areas is an example of one factor that might produce a different pattern of responses for another area. Notwithstanding the above limitations, the data should still be significant inasmuch as a large proportion of

TABLE VII

RESPONSES CATEGORIZED BY PARENTAL STATUS, FORMAL EDUCATION, AND AGE OF RESPONDENTS WHEN ASKED TO STATE ONE GOOD FEATURE OF ALBERTA'S EDUCATIONAL PROGRAM (N = 479)

If you had to state one feature of our schools that you think is good, what would this be?	Parental Status					Formal Education						Age of Respondents					
		Married															
Response Category	Single	N. C.	C. G. U.	P. S. O.	C. A. S.	ELEM.	H. S.	UNIV.	TECH.	OTHER	NO. ST.	Under 21	21-34	35-50	Over 50	Not Stated	Total
Don't know or couldn't say	33	31	22	35	60	61	89	13	11	4	3	13	74	60	31	3	181
Nothing good	2	2	3	6	5	1	10	5	--	1	1	1	9	5	2	1	18
All round good job	14	16	18	3	36	27	49	6	5	--	--	3	26	33	24	1	87
Teachers and teaching methods	5	5	2	1	33	9	23	12	2	--	--	1	19	23	2	1	46
Curriculum	7	8	3	4		3	20	10	2	2	--	3	12	16	6	--	37
Organization of school system	3	2	5	3	6	1	14	2	2	--	--	2	7	9	1	--	19
Physical facilities	2	1	7	5	11	4	19	1	2	--	--	1	7	14	4	--	26
Extra-curricular activities	1	3	1	4	3	2	8	1	1	--	--	--	6	5	1	--	12
Provision for individual differences	1	1	1	1	6	1	7	2	--	--	--	--	2	8	--	--	10
Miscellaneous replies	9	5	3	3	16	7	18	7	2	1	1	7	12	12	5	--	36
Response not indicated	1	--	1	1	4	3	2	1	--	--	1	--	3	1	3	--	7
Total	78	74	66	66	195	119	259	60	27	8	6	31	177	186	79	6	479

TABLE VIII

RESPONSES CATEGORIZED BY PARENTAL STATUS, FORMAL EDUCATION, AND AGE OF RESPONDENTS WHEN ASKED TO MAKE ONE SUGGESTION THAT WOULD IMPROVE ALBERTA'S EDUCATIONAL PROGRAM (N = 479)

If you had one suggestion to make to improve the calibre of our Educational System, what would this be?	Parental Status					Formal Education						Age of Respondent					
		Married															
Response Category	Single	N. C.	C. G. U.	P. S. O.	C. A. S.	ELEM.	H. S.	UNIV.	TECH.	OTHER	NO. ST.	Under 21	21-34	34-50	Over 50	Not Stated	Total
Don't know, couldn't say	30	28	25	25	47	52	85	5	6	4	3	10	64	46	33	2	155
Satisfied with present schools	1	4	2	--	9	7	5	2	1	1	--	--	5	7	4	--	16
Consideration of individual differences	2	2	5	2	11	4	11	5	2	--	--	2	10	7	3	--	22
Discipline (more or less of same)	7	11	4	5	16	12	25	5	1	--	--	3	15	17	8	--	43
Homework (more or less of same)	1	1	2	3	6	3	6	1	2	1	--	1	2	8	2	--	13
Teachers: Qualifications, pay, methods	7	8	8	6	29	11	31	14	2	--	--	4	17	30	5	2	58
Degree of emphasis of present curriculum	4	6	2	4	22	4	24	8	1	1	--	2	13	15	7	1	38
Extending curriculum	14	2	3	7	9	5	20	4	5	--	1	6	14	8	7	--	35
Restricting curriculum	2	3	7	4	19	7	19	6	3	--	--	--	12	18	5	--	35
Standards (lower, raise)	1	7	4	5	9	4	14	6	2	--	--	--	11	12	3	--	26
Administrative arrangements	6	--	1	2	9	3	11	2	2	--	--	2	7	8	--	1	18
Moral attitudes and values	1	2	1	2	7	5	6	1	--	1	--	--	4	8	1	--	13
Response not indicated	2	--	2	1	2	2	2	1	--	--	2	1	3	2	1	--	7
Total	78	74	66	66	195	119	259	60	27	8	6	31	177	186	79	6	479

the provincial population do live in the cities of Alberta.

Summary of Findings

There is one finding of this study that stands out above all others. This is the evident lack of interest of the sampled population in their schools. Sizeable proportions of the sampled group within the limitations of the survey method could neither criticise nor state one good feature of the schools nor make a single suggestion for their improvement. The establishment of a Royal Commission to study the status of education in Alberta has not created a marked demonstration of the typical citizen's interest in education even though the press and radio coverage of this Commission's activities has been considered generally to be excellent. When those recognizing the existence of this Commission were questioned further, less than a quarter of the total group surveyed were acquainted, even partially, with its aims and objectives. The same proportion expressed optimism about the ultimate worth of such a Commission.

This finding is all the more significant since any bias in sampling must have emanated from a group with a generally higher standard of living, access to a greater number and variety of communication outlets, and whose schools generally are staffed with teachers of higher than average qualifications and whose curriculum is much broader than generally found in the rural areas.

The study has brought out that the amount of formal education received by the respondent of the sample surveyed relates most closely to both knowledge of and willingness to criticise and suggest in educational matters. Those with the greatest amount of formal education demonstrate considerably more intelligent awareness than those with only an elementary education. While analysis of differences of responses in relation to age of the respondent or to the number of children in the family or the number that have or presently are attending or are about to attend school, are of such interest to warrant detailed statistical analysis of responses, only those relationships readily evident by inspection have been considered at this time.

It was not the purpose of this study to explain the reasons for the findings. However, they do not differ from those of other investigators. Of three recent Canadian studies (1, 2, 5), only that of Brimacombe (2) is similar to the present study. His findings are confirmed to a large extent by this study and suggests that no marked change in public attitudes has occurred in the intervening year and a half. Moreover, while those parents with children attending school showed a favorable attitude towards the schools in a similar pattern to that reported by Lindel (6) for St. Louis parents of school age children, the total population sampled in this study was found as a group to be unable or unwilling to commend their schools. Perhaps this is because the public, in general, has had no occasion to become concerned, positively or negatively, with education and prefers instead to leave such matters to the judgment of professional educators. It is only those citizens as a group having the equivalent or better formal education than these officials who openly question such matters. Perhaps the schools themselves, with the possible exception of home and school associations, have seen no reason to create citizen interest in education. It may be that what has been done in public relations by the schools and similar bodies, such as the Alberta Royal Commission on Education, has not been of a type suited to the general population but has been, as suggested by Hines and Grobman (4), more suited to the group with university or professional background. As stated earlier, this study has not attempted to answer these questions, yet they seem worthy of further investigation.

In brief, this study has shown that the average citizen is not inordinately interested in the public educational system and it may follow as a corollary that the reported outbursts and attacks on education are perhaps the product of a vocal and informed minority.

REFERENCES

1. A Survey Project of the Joint Committee on Public Attitudes Towards Our Schools, Nova Scotia Looks at Its Schools (Nova Scotia Department of Education, 1953).
2. Brimacombe, A. K. "The Construction of a Scale to Measure Adult Attitude Toward the Alberta Educational System," Alberta Journal of Educational Research, IV (June 1958), pp. 100-8.
3. Hand, Harold C. What People Think About Their Schools (New York: World Book Co., 1948).
4. Hines, Vynce A., and Grobman, Hulda, "What Parents Think About Their Schools and What They Should Know About Them," Bulletin of the National Association of Secondary School Principals (February 1957), pp. 15-25.
5. Larsen, O. P. What People Think About Their Schools, A Brochure and Summary, mimeographed, No. 4 (Alberta: County of Newell, 1954).
6. Lindel, Albert L. "What's Right With Them," The Phi Delta Kappan, XXXIV (June 1953), pp. 403-25.
7. The Public Looks at Education, Report No. 21 (Denver: National Opinion Center, 1944), p. 3.

8. Smith, Maurice M. "Comparative Social Attitudes," Journal of Educational Psychology, XXVIII (December 1937), pp. 681-93.
9. Spencer, Lyle M. What Los Angeles Educators Think of Their School System (Los Angeles: Los Angeles City Board of Education, 1953).
10. Stearns, Harry L. Community Relations and the Public Schools (Englewood Cliffs: Prentice-Hall, Inc., 1955).

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 1, September 1960)

SCHOLASTIC ACHIEVEMENT AT IOWA STATE COLLEGE ASSOCIATED WITH HIGH SCHOOL SIZE AND COURSE PATTERN

IRVIN T. LATHROP
Long Beach State College, Long Beach, California

IOWA STATE COLLEGE, as a state supported educational institution, admits all students who graduate from Iowa high schools, regardless of the academic attainment of the student while in high school. In recent years, Iowa State College has, however, been discouraging students whose high school grades average less than C from enrolling. However, since Iowa State College is a state supported institution, if the student insists, he will be admitted.

Students who enroll at Iowa State College represent a cross section of the high schools in the state of Iowa. A major portion of the secondary schools of Iowa are represented at Iowa State College. Graduates of high schools whose enrollments range from 7 to 2,350 attend Iowa State College. To ascertain if the students from small and large high schools have the same survival-attrition ratio and achieve equally well scholastically is one of the purposes of this study.

The other purpose of this study was to ascertain whether the various course patterns available in high schools prepared equally well for entrance to Iowa State College. All students who have graduated from Iowa high schools are eligible for admission to Iowa State College regardless of the courses they have taken in high school.

Sources of Information

The over-all study, of which this was a part, was concerned with 1516 non-transfer Iowa public high school graduates who entered Iowa State College in the freshman class of 1952. The four year high school enrollment for each of the classes was obtained from the Iowa Educational Directory. For purposes of IBM coding, ten size categories were established as follows: 1) less than 25; 2) 25-49; 3) 50-75; 4) 75-99, 5) 100-149; 6) 150-199; 7) 200-299; 8) 300-499; 9) 500-999; and 10) 1000 and over.

For purposes of analysis in this part of the study, the high school course patterns were divided into four classifications as follows: 1) mathematics-science; 2) college preparatory; 3) vocational; and 4) miscellaneous. For a student to be classified into the mathematics and science pattern, a total of at least 12 semesters of mathematics and science must have been completed. These 12 semesters could not include either biology or business arithmetic. The 12 semesters included two semesters of ninth grade algebra, two semesters of plane geometry, one semester of trigonometry, one semester of either advanced algebra or solid geometry, two semesters of general science, two semesters of chemistry and two semesters of physics.

In many instances the ninth grade was included in the junior high school. Consequently first year algebra and general science were not listed on the transcript. It was assumed, however, that to have been eligible to have taken advanced mathematics, chemistry and physics, the student would have credit for beginning algebra and general science. To be classified in the mathematics and science pattern, the student must have completed both high school chemistry and physics and six semesters of mathematics.

The college preparatory high school course pattern included eight semesters of English, eight semesters of social studies and history, two semesters of algebra, two semesters of plane geometry and enough other electives to complete a four year high school course. In many instances, the student classified as college preparatory had taken two or more semesters of a foreign language, such as Latin, French, or Spanish. Another requirement for eligibility in this classification was not more than two semesters of any type of vocational education course.

---

* Journal Paper No. J-3403 of the Iowa Agricultural and Home Economics Experiment Station, Ames, Iowa. Project No. 170.

To be classified into the vocational category, a student must have completed at least six semesters in one of the areas of home economics, agriculture, industrial education or business education in high school. The business curriculum usually included such courses as typing, shorthand, bookkeeping, office practice and in many instances a course entitled business law.

The remaining classification, miscellaneous, was reserved for cases which did not fit into any of the foregoing pattern of courses.

The transcripts were examined for compliance with a pattern of courses in the following order of preference:

1. Mathematics and science
2. College preparatory
3. Vocational
4. Miscellaneous

For example, whenever a student could be classified as having a mathematics and science pattern and one other, he was placed in the mathematics and science classification as it was thought that at a technical institution such as Iowa State College a mathematics and science background might be more beneficial than any other.

The control variables used for the analysis of covariance in this study were high school size, ACE scores and high school grade point average. The high schools from which the 1516 students graduated were classified into ten categories and coded "0" through "9" for purposes of IBM computation. High school grade point average and ACE scores were also classified into ten categories and coded "0" through "9". The statistical methods followed were those associated with regression and discriminant analyses.

Index of High School Size

The predominance of small high schools in the state of Iowa can be readily seen by an inspection of Table I. Although considerable change took place before and during the early 20's as a result of a wave of school district consolidation, this change actually resulted in more high school districts, rather than fewer, since this movement was in reality a consolidation of a few elementary school districts to form a new district which established a high school with a resulting small enrollment. In 1911-12, there were only 260 high school districts in the state of Iowa, but in 1931-32, as a result of the consolidation movement, there were 909 high school districts. Additional changes have taken place in recent years as a result of school district reorganization as shown in Table I. From the peak number of 909 high schools in 1931-32, the number of high schools has been gradually decreasing until in 1957-58 there were 745 high schools in the state of Iowa.

Although the number of districts maintaining high schools has fluctuated over the years, the state of Iowa has been, and still remains, characterized for its adherence to the practice of small units for high school instruction. The median high school size for the school year 1957-58 was 94 students. This median was, no doubt, the highest during the past 50 years although the typical high school had almost as great an enrollment during the first decade of the century. The smallest median was 57 students in 1921-22 for any of the years summarized in Table I.

The relationship between college achievement and size of high school is a major objective of this study. Philosophical considerations, based on some objective evidence and prevailing points of view in education, point to a curvilinear relationship between size of high school and college achievement regardless of the terms in which such achievement is evaluated.

To test whether the relationship was curvilinear, a test of significance of quadratic over linear regression was made by high school size and cumulative quality point average. Since this advantage resulted in significance at the 5% level, it was desirable to assume some type of curve. Because the size of school, for convenience, had been coded for purposes of IBM tabulation, it was expedient to assign indexes of high school size to the IBM code.

It was decided to find the relationship between the cumulative quality point average and the index of high school size based upon some root of the actual high school size. By a trial-and-error method, the eighth root of the actual enrollment was determined as the index of size to be used in rectifying this relationship. The eighth root of the actual enrollment with slight rounding of numbers may be expressed on a scale of "0" to "7" as shown in Table II. The process of rectification of the relationship seemed warranted since subsequent analysis of the advantage of quadratic over linear regression yielded a nonsignificant t-value of 1.07.

The eight intervals of high school size, from "0" to "7", were designated as the index of high school size. Throughout this study, all relationships between various characteristics of students and of high schools have been made with this index of high school size rather than with the actual high school enrollment.

No claim is made that the index of high school size, as here defined, is the best function of actual enrollment. It does represent a plan of rectifying the relationship between high school enrollment and college achievement which yields a far more sensitive test of significance than assuming that the interval of high school enrollment are not numerically ordered.

Criteria of Achievement

Of major concern to anyone undertaking a pre-

TABLE I

PERCENT OF HIGH SCHOOL DISTRICTS IN EACH SIZE CATEGORY FOR THE YEARS 1901 TO 1958 AND MEDIAN SIZE HIGH SCHOOL AND MEDIAN ENROLLMENT

Size	1901-02 N=176	1911-12 N=260	1921-22 N=852	1931-32 N=909	1941-42 N=897	1951-52 N=827	1957-58 N=745
Less than 25	1.0	3.1	16.4	5.8	6.0	3.8	1.5
25	13.6	11.5	27.6	20.5	20.5	24.3	18.1
50	25.0	26.9	20.3	26.6	26.6	24.5	19.6
75	16.6	16.2	12.3	16.4	13.7	14.1	14.2
100	25.0	21.2	10.2	13.2	15.2	14.7	19.7
150	8.5	7.3	4.0	4.5	4.8	6.2	7.8
200	4.6	7.3	4.6	5.5	4.8	4.5	8.1
300	5.1	4.2	2.8	4.3	5.0	5.2	6.3
500	0.6	1.9	1.3	1.8	2.1	1.3	4.7*
1000	0.0	0.4	0.5	1.4	1.3	1.4	
Median Size School	91	88	57	72	72	72	94
Median Enroll-ment	134	157	152	232	267	244	308

* No breakdown available for these categories for the years 1957-58

TABLE II

INDEX OF HIGH SCHOOL SIZE

Size of High School	IBM Code	Index of Size	Cumulative Quality Point Average
Less Than 25	0	0	1.25
25-49	1	1	2.01
50-74	2	2	2.08
75-149	3-4	3	2.06
150-299	5-6	4	2.09
300-499	7	5	2.04
500-999	8	6	2.21
1000 and over	9	7	2.25
All Schools			2.12

diction study is the selection of the criterion. The criterion employed should be as valid and useful as possible, yet readily obtainable. For purposes of the present study, four criteria were employed.

The first criterion was survival-attrition tendency at the end of the first quarter of college. Survival, as here defined, was the tendency to enroll more than one quarter as a student. No attempt was made in this study to ascertain the causes of freshman attrition. Such causes presumably might include: 1) lack of scholastic ability; 2) lack of motivation for academic proficiency; 3) inadequate counseling; 4) change in economic capacity; 5) illness or death; 6) marriage; 7) transfer to other universities; 8) failure to adjust to a college environment; and 9) many other causes or combinations of causes.

Included in the survival group are freshmen 1) who might have been denied registration at the beginning of the second quarter by action of the scholarship committee, but who at some subsequent date did re-enroll at Iowa State College and 2) who had changed educational or occupational objectives as indicated by a change in curriculum at Iowa State College.

Of the 1516 freshmen included in the total study, 1384, or 91.3% survived more than one quarter. The attrition at the end of the first quarter does not seem unduly excessive. Although no evidence had been assembled in this study to indicate the magnitude of changes from a curriculum in one division to a curriculum in another division, the attrition percentage might have been considerably higher if attrition would have been defined in terms of the first chosen curriculum.

The other dichotomized criterion was concerned with graduation-attrition tendency at the end of a five year period. Graduation, as here defined, was the tendency to graduate and obtain a degree. Again, no attempt was made to ascertain the causes of student attrition. The causes might presumably include those postulated for attrition at the end of the first quarter with the added possibility of military service. There may be other causes or combinations of causes which contribute to student attrition which are unidentifiable.

Included in the graduation group were those students who were still enrolled at the end of a five year period. It was assumed that these students would eventually earn a degree. In many instances these students, who were still enrolled, were young men who had their education interrupted by military service and upon discharge from the military had re-enrolled at Iowa State College. Students who had changed from a curriculum in one division to a curriculum in another division during the five years, but who earned a degree from Iowa State College, were classified as graduating.

Of the 1516 freshmen included in the total study, 680 or 44.9% graduated. Although no evidence has been assembled in this study to indicate the magnitude of changes from one curriculum to another, the attrition percentage, no doubt, would be considerably higher if attrition again had been defined in terms of the first chosen curriculum.

The other criteria employed were quality point average at the end of one quarter of college and quality point average at the end of a five year period. For each credit earned, a student receives quality points, according to the mark attained, as follows: A, 4 points; B, 3 points; C, 2 points; D, 1 point; F, O points. The quality point average for each student at the end of the first quarter was obtained by dividing the number of quality points attained by the number of credits attempted.

The quality point average at the end of a five year period, designated as cumulative quality point average, was obtained by dividing the total number of quality points earned for the five years by the number of credits attempted for the five year period.

One of the major causes of student attrition is an unsatisfactory achievement as reflected by low or failing quality point averages. Since one of the causes of attrition is unsatisfactory achievement as indicated by a low quality point average, there should be a substantial correlation between survival-attrition tendency and quality point average. A biserial correlation of 0.325 was found between survival-attrition tendency and first quarter quality point average for students included in this study. This correlation was of sufficient magnitude to be significantly different from zero. However, it was not as high as one might expect if the only cause of student attrition was low quality point average.

The biserial correlation between graduation-attrition tendency and cumulative quality point average at the end of a five year period was 0.372 which was slightly higher than the correlation found for the first quarter. The magnitude of these correlations would seem to indicate that, to a large degree, the criteria are somewhat different measures of college achievement.

It should be pointed out, however, that these correlations may be smaller than would normally be anticipated because the students who transferred to other universities were placed in the attrition category. Another factor contributing to the lowering of the correlations is the unreliability of the four criteria.

The purpose of secondary education has been stated in terms of developing the person into an effective member of a democratic society. The criteria here chosen are useful in evaluating the effect of various high school course patterns and size of high school in only one respect, that of preparation for study at Iowa State College. It is not the purpose of this study to evaluate the various high school course patterns in terms of the objectives for which they may have been designed.

The four criteria, however, constitute legitimate standards of evaluating the size of high school and high school pattern from the viewpoint of potential students at Iowa State College. These students represent approximately one in every twelve graduates of the public high schools in Iowa.

Findings

To ascertain whether the size of the high school from which a student graduated had any significant effect on achievement at Iowa State College, a sample of 180 cases was drawn at random from the 1516 included in the total study. The sample of 180 cases was used in an effort to minimize the effects of the confounding of high school size and high school course pattern. As can be seen by Table III, agriculture appears to be a course pattern offered predominantly in the middle size high schools, whereas industrial education, to the degree necessary to be included in this study, is offered mostly in the larger high schools.

To obtain enough cases from each of the course patterns for analysis, high school size was classified into three categories: under 100, 100-499, and 500 or greater enrollment. To further simplify the analysis, the agriculture, home economics, industrial education and business course patterns were combined into one pattern, classified as vocational. Thus, for the analyses, the four high school course patterns of mathematics and science, college preparatory, vocational and miscellaneous were used.

Disproportionality existed among high school course patterns and high school size as shown by Table IV. There were only 15 of the 1516 students included in this study who entered Iowa State College from high schools of under 100 with the mathematics and science course pattern, whereas there were 197 students who entered Iowa State College from high schools of 500 or over with this same course pattern. However, the largest number of students who entered Iowa State College with the miscellaneous course pattern came from smaller size schools. The confounding effect of high school course pattern and high school size is indicated by a chi square of 242 which is significant far beyond the 1% level.

In order to eliminate this confounding effect between high school size and high school course pattern, the number of cases in each cell of Table IV was randomly sampled down to 15, the number appearing in the smallest cell. In each group of 15 students, there was a proportional number drawn at random from each of the divisions of the college, making a total of 180 students.

Index of high school size, ACE score and high school grade point average were used as variables. A correlation coefficient computed between index of high school grade point average on the entire group of 1516 students yielded a value of -0.184 which was significantly different from zero. Thus, it would appear from this correlation that high school grades are partially a function of high school size. It is a negative relationship, meaning that students from smaller high schools tend to have higher high school grade point averages. A correlation coefficient computed between the index of high school size and ACE scores yielded a value of 0.055 which was not significant with 1516 students.

The correlation between the index of high school size and the four criteria are shown in Table V. The correlations are shown without control, with control on ACE scores, with control on high school average and with control on both ACE scores and high school grade point average.

The coefficients between the index of high school size and the two dichotomized criteria were not significant, regardless of what other variable or combination of variables were controlled. However, with the two criteria of quality point averages, the correlations were significant when controlling on high school grade point averages and when controlling on both ACE scores and high school grade point averages.

It appears that whenever high school grade point average was used as a controlling variable it raised the relationship of the index of high school size the the two criteria of quality point averages to a magnitude which was significant. Thus, it would appear that whenever first quarter or cumulative quality point averages were used as criteria, controlling on high school grade point average might possibly introduce some bias. Any bias which high school grade point average might introduce as a control variable should not, however, preclude its use in a prediction equation. The techniques of regression and discriminant analyses were employed in this study. To find whether high school size had significant effect on achievement at Iowa State College, the index of high school size was eliminated as a prediction variable and the loss tested for significance. The regression equations and the coefficients of correlation here reported are those obtained within the high school pattern of courses.

Because of the possibility of bias being introduced by using both the index of high school size and high school grade point average as control variables, the index of high school size was eliminated as a control for subsequent analyses. Analysis of covariance was employed to test whether the high school pattern of courses had a significant effect on college achievement.

Survival-Attrition Tendency

The first criterion used to measure student achievement was survival-attrition tendency at the end of the first quarter of college. The percent-

TABLE III

PATTERN OF COURSES OF STUDENTS AT IOWA STATE COLLEGE BY HIGH SCHOOL SIZE

Size	Math. - Science	Ag.	Home Ec.	College Prep.	Industrial Ed.	Business	Misc.	Total
Less Than 25	0	0	0	0	1	1	1	3
25	3	1	2	13	0	22	27	68
50	6	13	7	30	5	34	48	143
75	6	14	6	39	0	32	40	137
100	18	35	9	57	7	28	56	210
150	16	16	4	32	3	11	12	96
200	38	26	8	39	8	12	9	140
300	73	30	5	67	16	16	21	228
500	118	7	1	86	48	23	11	294
1000+	79	0	8	69	24	5	12	197
Total	357	145	50	432	112	184	236	1516

TABLE IV

PATTERN OF COURSES OF STUDENTS AT IOWA STATE BY THREE HIGH SCHOOL SIZE CATEGORIES

Size	Math. -Science	College Prep.	Vocational	Miscellaneous	Total
Under 100	15	82	138	116	351
100-499	145	195	236	98	674
500 or more	197	155	116	23	490
Total	357	432	490	237	1516

TABLE V

COEFFICIENTS OF CORRELATION BETWEEN THE INDEX OF HIGH SCHOOL SIZE AND THE FOUR CRITERIA

Criterion	No Control	Within Pattern Correlation with Control		
		ACE	HSA	Both
First Quarter Quality Point Average	0.158	0.015	0.232	0.181
Cumulative Quality Point Average	0.090	0.006	0.197	0.152
Survival-Attrition Tendency	0.180	0.018	0.074	0.058
Graduation-Attrition Tendency	0.079	0.012	0.093	0.052

age of students surviving past the first quarter for the four high school course patterns were:

Mathematics-science	95.6%
College preparatory	97.8%
Vocational	91.1%
Miscellaneous	82.2%

An analysis of the course patterns yielded an F-value of 40.8 which was significant well beyond the 1% level. Thus, from an analysis in which high school size, ACE score and high school grade point average were ignored, it appeared that the high school pattern of courses had an effect on survival-attrition tendency at the end of the first quarter of college.

A discriminant equation to predict the chances in 100 of any given student surviving past the first quarter was developed. Even though the index of high school size contributed but little to prediction, it was included along with the ACE score and the high school grade point average for predicting survival.

Because the analysis of variance of survival-attrition tendency among high school course patterns does not take into account differences in high school size, ACE score or high school average, the more refined technique of analysis of covariance was used to control on these variables. The analysis of covariance among high school course patterns was computed and an F-value of 39.7 was obtained. This F-value was significant well beyond the 1% level, indicating that with control on high school size, ACE score and high school grade point average, the high school pattern of courses had an effect on survival attrition tendency for the first quarter of college.

The index of high school size makes little contribution to either prediction or control when the effects of the confounding with course pattern are eliminated. For this reason, analysis of covariance was computed without the use of the high school size as a control variable. This F-value was not appreciably different from that attained when three control variables were used.

When using the criterion of survival-attrition tendency at the end of the first quarter, the index of high school size has little value for predicting survival or as a control variable for the analysis of covariance among high school course patterns. However, the chances of a student surviving past the first quarter at Iowa State College varies with the high school pattern of courses.

First Quarter Quality Point Average

The other criterion employed for evaluating student progress at the end of the first quarter was quality point average. The mean first quarter quality point averages for the four high school course patterns were:

Mathematics-science	2.53
College Preparatory	2.41
Vocational	2.17
Miscellaneous	1.60

An analysis of variance of the course patterns yielded an F-value of 11.0 which was significant beyond the 1% level. Again, when high school size, ACE score and high school grade point average were ignored, the pattern of courses a student completed in high school had an influence on quality point average at the end of the first quarter.

A regression equation was developed for predicting a first quarter quality point average for any given student using the three variables of high school size, ACE score and high school grade point average. The loss due to the elimination of the index of high school size was computed and an F-value of 6.0 was obtained which was significant at the 5% level. A partial coefficient of correlation between the index of high school size and first quarter quality point average, holding constant ACE score and high school grade point average, was 0.181. However, both the F-value and the correlation may have been unduly high due to the possible bias suggested by a significant correlation between the index of high school size and the high school grade point average. Because of the possibility of this bias existing, another regression equation was developed without the index of high school size.

The more refined technique of analysis of covariance was used to control on high school size, ACE score and high school grade point average. The analysis of covariance of first quarter quality point average among high school course patterns yielded a nonsignificant F-value. Evidence from this group of students failed to reveal any difference among high school course patterns when evaluated in terms of first quarter quality point average for students who are statistically equated on high school size, ACE score and high school grade point average.

The high school pattern of courses had little effect on first quarter quality point average when either of the two variables of high school size or high school grade point average were used as control variables. However, when ACE score alone was used as a control variable, the high school pattern of courses had an effect on the quality point average a student earns at the end of the first quarter of college.

Too much confidence should not be placed in student achievement, either survival-attrition tendency or quality point average, when evaluated at the end of one quarter of college. It would seem reasonable to expect that the reliability of the criteria would be increased as evidence is accumulated over a longer period of time.

Graduation Attrition Tendency

Survival past the first quarter is an indication of eventual graduation. However, it is no guarantee of graduation as evidenced by the fact that less than 55% of the students who survived the first quarter evantually graduated. The percentage of students graduating from Iowa State College for each of the four high school course patterns was:

Mathematics-science	57.8%
College preparatory	62.2%
Vocational	51.1%
Miscellaneous	26.7%

An analysis of variance of the course patterns yielded an F-value of 5.4 which was significant at the 1% level. From an analysis in which high school size, ACE score and high school grade point average have been ignored, it appears that the high school pattern of courses had an influence on graduation-attrition tendency at Iowa State College. Equations were developed utilizing: 1) the high school size, in addition to the ACE score and high school grade point average and 2) the ACE score and high school grade point average alone.

The index of high school size did not contribute much as a prediction variable as none of the partial coefficients of correlation were significant. A multiple correlation between the index of high school size and ACE score with graduation-attrition tendency was 0.3155, whereas the correlation between ACE score alone and graduation-attrition tendency was 0.3153.

The amount that the sigma value would change from the largest to the smallest size high school is only 0.028 which would seldom change the graduation percentage more than two percent. This slight difference would be of no practical use in the counseling of students.

Because analysis of variance does not take into account differences in student ability, the more refined technique of analysis of covariance was used to control on high school size, ACE score and high school grade point average. When these three variables were used as controls, the high school pattern of courses had no significant influence on graduation-attrition tendency. Another analysis of covariance was computed using ACE score alone as a control. When ACE score alone was used as a control, the high school pattern of courses affected the graduation-attrition tendency at Iowa State College.

When using the ctiterion of graduation-attrition tendency at the end of a five year period, the index of high school size has little value for predicting graduation or as a control variable for the analysis of covariance among high school course patterns. If ACE score alone is held constant, the chances of a student surviving the first quarter or graduating from Iowa State College varies with the course pattern which he completed in high school.

Cumulative Quality Point Average

The other criterion employed for evaluating student progress at the end of a five year period was cumulative quality point average. The mean cumulative quality point averages for the four high school course patterns were:

Mathematics-science	2.39
College preparatory	2.33
Vocational	2.29
Miscellaneous	1.63

An analysis of variance of the course patterns yielded an F-value of 10.41 which was significant at the 1% level. When high school size, ACE score and high school grade point average were not considered, the high school pattern of courses had an influence on cumulative quality point average at Iowa State College.

Again, regression equation were developed for predicting an individual student's cumulative quality point average. One regression equation was developed using the three variables of high school size, ACE score and high school grade point average. Other equations were developed without the use of high school size.

Analysis of covariance among high school course patterns was used to control on the index of high school size, ACE score and high school grade point average. With these three variables controlled, the analysis yielded a nonsignificant F-value. Another analysis of covariance was computed using ACE score alone as a control variable. This analysis yielded an F-value which was significant at the 5% level. It appeared that the quality point average which a student earned varied with his high school pattern of courses when ACE score alone was held constant.

More confidence can probably be placed in student achievement, either graduation-attrition tendency or quality point average, when evaluated at the end of a five year period. The reliability of the criteria was undoubtedly increased by evaluating at a later date in the student's academic career.

The use of high school average as a control variable in the analysis of covariance among high school course patterns is open to some limitations. To be entirely satisfactory, the high school grade point average, for the same quality of performance, should be identical regardless of the size of high school. No direct evidence was available to substantiate or refute this principle. Some indirect evidence suggests that the smaller high schools tend to give higher grades than the larger high schools as indicated by a coefficient of correlation of -0.184.

It does not necessarily follow from this correlation that the standard of grading varies with size of school. It might be that Iowa State College is drawing a better caliber of students from small high schools than from large high schools. This latter interpretation does not seem tenable, however, when we note the nonsignificant correlation of 0.055 between ACE score and size of high school.

TABLE VI

PERCENTAGE CONTRIBUTIONS TO VARIANCE OF COLLEGE ACHIEVEMENT BY PREDICTION VARIABLES

Criterion	Index of High School Size	High School Course Pattern	ACE Score	High School Grade Point Average	Unexplained
Survival-Attrition	0.2	65.4	1.4	2.9	30.1
First Quarter Quality Point	0.8	0.5	12.9	33.0	52.8
Graduation-Attrition	0.1	3.0	6.4	8.8	81.7
Cumulative Quality Point	0.4	1.0	9.9	32.2	56.5

College Achievement and Size of High School

Little evidence can be found concerning the relationship of size of high school to college achievement except in a sample in which the confounding effect with high school course pattern has been eliminated. This has been done in the foregoing analyses by reducing the number of students in the study from 1516 to 180.

In this sample, the contributions to the variance of college achievement by size of high school has been computed and is shown in Table VI. It can be seen that in no case does the percentage contribution of high school size reach one percent when index of high school size, ACE score, high school average and high school course pattern are considered. The negative relationship between the index of high school size and high school grade point average sugeests that the percentage contributions of the index of high school size, small though they be, are inflated. If the contribution of high school size is determined when only ACE score and high school course pattern are considered, the contribution of the index of high school size in no case reaches one-tenth of one percent. This suggests that the size of high school, in itself, is almost unrelated to college achievement, although it doubtless is a considerable factor in determining the type of course pattern available in the high school.

Summary

Because disproportionality existed between the size of the high school from which a student graduated and the high school course pattern, a sample of 180 cases was drawn at random from the total sample of 1516 students. When using the criterion of survival-attrition tendency, the size of the high school had little value as a prediction variable or as a control variable in the analysis of covariance. However, the chances of a student surviving the first quarter varies with the high school pattern of courses which he completed.

When achievement is evaluated in terms of quality point average at the end of the first quarter, the size of high school has little value as a prediction variable or as a control variable in the analysis of covariance. When ACE score alone was held constant, the first quarter quality point average varied with the pattern of high school courses completed.

When graduation-attrition tendency is used as a criterion to evaluate student achievement, the high school size has little value for predicting graduation or as a control variable when using analysis of covariance. When ACE score alone was held constant, the chances of a student graduating varied with the high school course pattern.

When cumulative quality point average is used to evaluate student achievement, the size of the high school may be disregarded. When ACE score alone was held constant, the cumulative quality point average varied with the high school pattern of courses.

From analysis on a proportional sample, it appears that the size of the high school from which a student graduated has little influence on achievement at Iowa State College when the confounding effect high school size and course pattern is eliminated. The pattern of courses a student completes in high school greatly influences his achievement at Iowa State College. It seems reasonable to conclude that, if the small high schools could offer the diversity of courses that the larger high schools offer, there would be no difference in the college achievement of graduates from the large and small high schools.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 1, September 1960)

A STUDY OF UNDERGRADUATE RECORDS OF GRADUATES FROM HUNTER COLLEGE WHO LATER EARNED DOCTORATES

EDITH B. CHASE
Hunter College, New York City

IN THE FALL of 1957, the Office of Scientific Personnel of the National Academy of Sciences and the National Research Council sent to each college a list of its graduates who had been awarded the doctorate by other institutions between the years 1936 and 1956. With this list were given for each individual the date of the baccalaureate and the field and date of the doctoral degree. This information seemed to offer an excellent opportunity to ascertain for a complete group of students who came from a single undergraduate body and who later earned the doctorate, the degree to which the quality of undergraduate work indicated potential doctoral material, and to determine any differences between the records of prospective graduate students in different disciplines.

Of the 304 names on the list, ten were omitted for various reasons. Of the individuals included in this study, all but 13 were women, for, until veterans were admitted in 1946, male students were limited to a very small number in the evening session. From the undergraduate record of each student named in the list, the following data were obtained: Cumulative index (each credit of A counting 4 points, B three points, C two points, D one point, and F no points); Index of first semester; Index of courses in the field of the future doctorate or of the major, if related to the doctorate (i.e., Chemistry major for doctorate in Biochemistry); High school average (where recorded); Honors (Junior Phi Beta Kappa, Senior Phi Beta Kappa, summa cum laude, magna cum laude, and cum laude); Country of birth of student and parents where recorded.

The students at this college differed in some respects from students in most other colleges, since they attended a free tuition non-residential college in the city of New York. That recent immigrants to this country took advantage of the opportunity for a free education for their children is indicated by the following facts. Of 270 students for whom the country of birth was recorded, 42, or 16%, were foreign born. Of 114 students entering between 1923 and 1939 when the birthplace of the parents was recorded, it was found that both the parents of 63, or 55%, were foreign born, one of the parents of 20 students, or 18%, was foreign born; while 31 students, or 27%, had both parents native born.

Since it seemed desirable to obtain some indication of the way in which the records of the students who subsequently earned the doctorate compared with records of all the students for the same period, a random sample of 294 of the latter was chosen. To offset any changes in grading standards over this period of years, the same number of individuals was picked from each graduating class as was included in the original list. The records were chosen by applying numbers from a list of random figures (2) to the numbered alphabetized list of graduates for each semester. The data taken from each record included in the random sample were limited to the cumulative index, the index of the first term, and honors.

Prospective Doctors Versus the Average Student

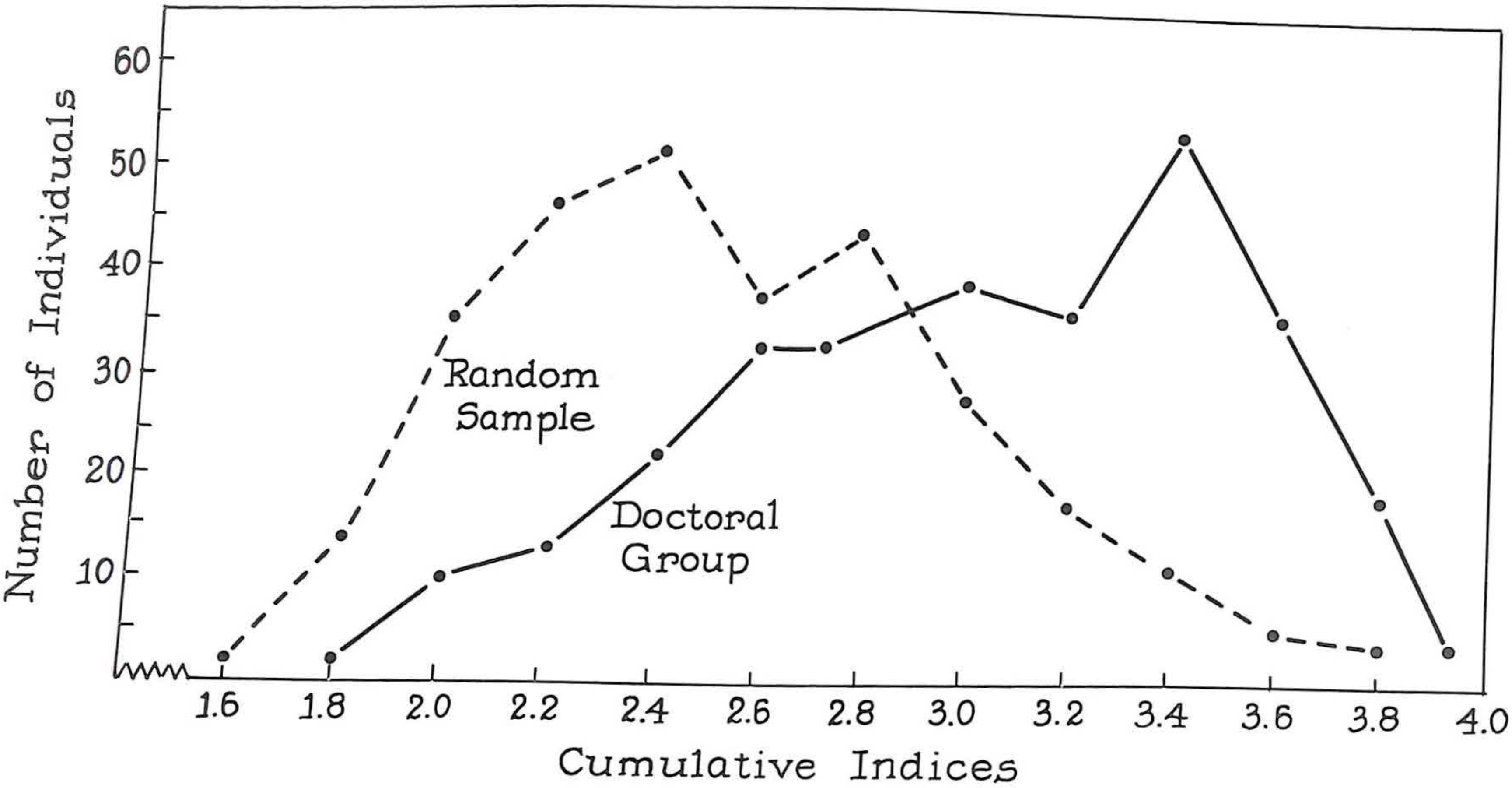
In Figure 1 are shown the distributions of the cumulative indices of the undergraduate records of the future doctors and of the individuals of the random sample. It is obvious that, whereas the indices of the random sample fall into a curve which, like most grade curves, is skewed to the right, the indices of the doctors fall into a curve which is skewed to the left. The Pearsonian measure of skewness for the sample is -0.248 and for the doctors +0.302.[1]* The difference of 0.48 between the mean of 3.04 $\pm$ 0.496 for the doctors and 2.56 $\pm$ 0.466 for the sample is highly significant with a t value of 12.3 for 294 cases of each. The probability of this difference being insignificant is less

---

* Footnotes will be found at the end of the article.

FIGURE 1

DISTRIBUTION OF CUMULATIVE INDICES OF STUDENTS WHO LATER EARNED THE DOCTORATE AND OF STUDENTS IN THE RANDOM SAMPLE



than 0.001.

In Figure 2, the distributions of the indices for the first term in college are shown for the doctors and for the sample. The curves here are plotted with intervals of 0.3 in indices instead of 0.2 as was used in Figure 1, because, with the wider range of the indices of the first term and the tendency for sets of grades earned in a single semester to fall into patterns, the interval of 0.3 was found to give a clearer picture of the distribution. Here both curves are skewed to the left, due to the very low grades of some beginning students, but there are more individuals with indices above 3.0 among the doctors than among students in the random sample. The differences between the mean of 2.81 ± 0.69 for the 294 doctors and 2.30 + 0.66 for the sample is 0.51 with a t value of 9.0 and the probability of a random occurrance less than 0.001.

To determine the degree of correlation between the cumulative indices for the entire undergraduate period and the first term grades, the coefficient of correlation was calculated for each group. For the sample the coefficient of correlation was 0.71 with a t value of 17.41, the agreement of the two indices having a high significance with a probability factor less than 0.001. For the doctors, the coefficient of correlation was considerably higher, 0.86 with a t value of 29.05 and a probability far below 0.001.[2]

The coefficient of correlation was also determined for the cumulative indices of the doctors and their high school averages where the latter were available. These data are not particularly reliable because admission to the college during some periods was based on the weighted average of high school grades and at other times on the weighted average of the Regents' Examinations of New York State which were highly variable. Nevertheless, the coefficient of correlation of the 266 records included was 0.66 with a t value of 14.11 and a probability of less than 0.001.

It is seen therefore, as was expected, that the undergraduate grades of students who later earned the doctorate are significantly higher than those of the average student, even in grades earned in the first term and at the high school level. The surprising fact, however, is the occurrence of low grades at all three levels of a few students who later obtained the doctorate. Among the 294 students in this group, four were dropped for poor scholarship at some time during their undergraduate years, and later earned rematriculation by attending courses given in the evening session as non-matriculated students. Two of these students obtained the doctorate in Education, one in Chemistry and one in Psychology.

Value of First Term Index as an Indicator of Academic Potential

There is currently in many colleges an interest in providing an enriched and stimulating curriculum for students with potentialities for superior work and advanced study. To this end, it is highly desirable to be able to identify a large majority of these students near the beginning of their college course. For this the first term indices and cumulative indices of the doctoral group were compared to determine with what reliance success as lower freshmen is a predictor of success in the total college curriculum.

In Table I is shown a comparison between the first term indices and the cumulative indices grouped in intervals of 0.5. Of 50 students with a first term index of 3.5 or above, 39, or 78%, finished with a cumulative index in the same range, and the remaining 22% had indices above 3.0. Of 90 with a first term index between 3.0 and 3.4, 92% completed their college work with an index of at least 3.0, with only 8% in the group below 3.0. Of 89 students with a first term index below 2.5, only two reached a 3.0 cumulative index.

In Table II is a similar comparison of the first term indices and cumulative indices of the random sample. The lower academic success of the 294 persons composing the random sample as contrasted with the similar number shown in Table I is apparent. If, however, we examine the most superior among the random sample, it is evident that the conclusions reached above hold. Five of nine students with a first term index above 3.5 had a cumulative index above 3.5, and the four others had cumulative indices above 3.0; of those having a first term index of 3.0 to 3.4, 27 (66%) completed their undergraduate work with an index above 3.0.

The outcome of this survey indicates that superior students in almost all cases are superior in their first term, especially in the more outstanding group, and a selection of students with a first term index of 3.5 or above would yield a group with a high percentage of prospective worth as doctoral candidates.

Comparison of Grades of the Doctoral Group in Different Fields

For a comparison of the grades of the doctoral group in the different divisions of subjects, the undergraduate records were separated into the main fields in which the different disciplines were grouped in the report by the Office of Scientific Personnel.[3] Included within the physical sciences are chemistry, geology, mathematics and physics; within the biological sciences, biology, biochem-

FIGURE 2

DISTRIBUTION OF FIRST TERM INDICES OF STUDENTS WHO LATER EARNED THE DOCTORATE AND OF STUDENTS IN THE RANDOM SAMPLE

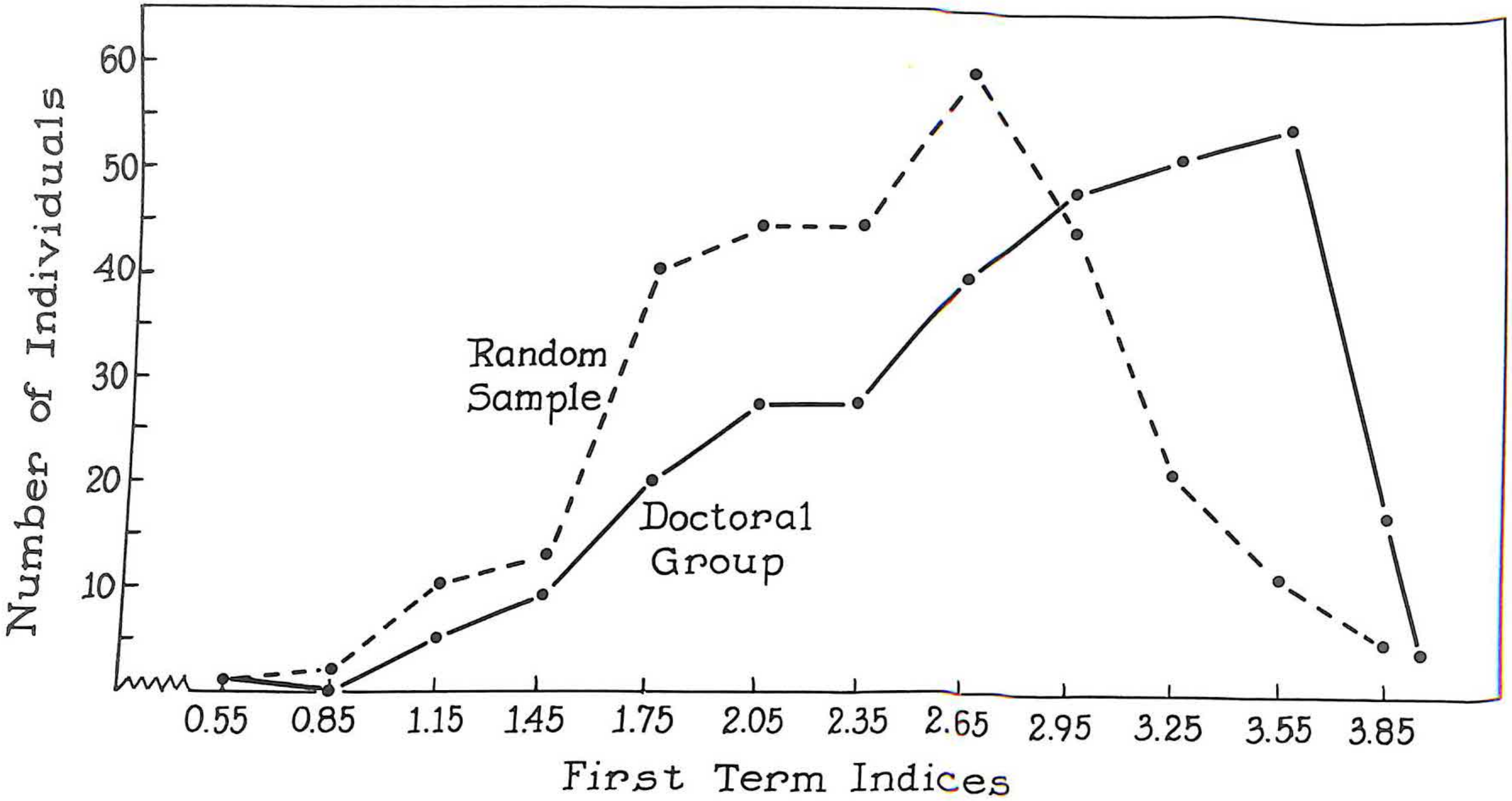


TABLE I

COMPARISON OF FIRST TERM INDICES AND CUMULATIVE INDICES OF 294 STUDENTS WHO LATER EARNED THE DOCTORATE

Cumulative Indices	First Term Indices						
	0.4-0.9	1.0-1.4	1.5-1.9	2.0-2.4	2.5-2.9	3.0-3.4	3.5-4.0
1.6-1.9	0	1	1	2	0	0	0
2.0-2.4	1	11	13	13	5	0	0
2.5-2.9	0	0	12	33	31	7	0
3.0-3.4	0	0	0	2	28	64	11
3.5-3.9	0	0	0	0	1	19	39

TABLE II

COMPARISON OF FIRST TERM INDICES AND CUMULATIVE INDICES OF 294 STUDENTS IN THE RANDOM SAMPLE

Cumulative Indices	First Term Indices						
	0.4-0.9	1.0-1.4	1.5-1.9	2.0-2.4	2.5-2.9	3.0-3.4	3.5-4.0
1.6-1.9	1	8	12	4	2	0	0
2.0-2.4	4	17	39	50	10	0	0
2.5-2.9	0	1	11	23	45	14	0
3.0-3.4	0	0	1	1	15	21	4
3.5-3.9	0	0	0	0	0	6	5

istry, biophysics, physiology, pharmacology and premedical work; within the arts and humanities, fine and applied arts, business, classical languages and literature, English and American literature, humanities, home economics, library and archival sciences, law and jurisprudence, linguistics, modern foreign languages and literature, music, philosophy and speech; within the social sciences, anthropology, archeology, economics, geography, history, political science and sociology. Psychology and Education were treated as separate fields.

In Table III, the average cumulative index for the students in each of these fields is given in the second column. With the exception of Education, the differences between the groups are small. By computing the estimated variance (1) of these groups, we obtain an F value of 1.3, with a probability greater than 0.10, indicating that these differences are not significant.

The average for those who obtained the doctorate in Education[3], however, differed widely from the averages of all the other groups. The curves of distribution of the indices for these two groups have been plotted in Figure 3, along with the curve for the random sample. The curve for the doctors in Education is bimodal; the mode of the majority, 2.4, coincides with the mode of the random sample, and the mode of the minorty, 3.4, coincides with the mode of the non-education doctors. The difference of 0.47 between the means of the education and non-education groups is highly significant with a t value of 7.1 and a probability less than 0.001. The difference of 0.15 between the mean of the sample and the mean of the education group is barely significant with a t value of 2.0 and a probability of 0.05. The distribution of indices in all other fields fell into curves similar to that of the non-education group.

The explanation of the bimodal curve in the education group can only be surmised. It cannot be explained by the assumption that superior students obtained the degree soon after graduation, while mediocre students teaching at the secondary or elementary level started graduate work later or took more time to complete it, since the average index was 2.63 for 13 obtaining the degree within ten years, 2.68 for 40 persons taking between 11 and 20 years, and 2.70 for 11 taking between 21 and 33 years.

Although there are no significant differences between the average indices of the fields, with the exception of Education, the picture is somewhat different when the averages of separate subjects are compared. In Table IV are listed the average indices of all subjects, except Education, in which there were three or more individuals. The means vary from 3.61 for Mathematics to 2.81 for Sociology. By computing the estimated variance for these groups, we get an F value of 1.91, with a probability less than 0.05. In the fields defined above, such widely divergent subjects as Mathematics and Geology are both grouped together in the field of physical sciences, a fact which tends to obscure the differences between the subjects.

Comparison of Cumulative Indices Versus Indices Obtained in Field of Doctorate

In order to determine to what extent special abilities in particular subjects may be obscured by the application of the cumulative index to cover all undergraduate work, the indices earned in the subject of the doctorate were determined, or in the major if related to the doctorate; i.e., in an undergraduate Chemistry major for a person obtaining the doctorate in Biochemistry. Whenever less than six credits were taken in the subject of the doctorate, individuals (29) were omitted from this study. In the third and fourth columns in Table III are shown the number of students in each field and the average index in the doctoral subject, and in the last column are given the differences between the cumulative index and the index in the subject of the doctorate. The index in all cases in the field of Education is for courses taken as a minor or as optionals, since a major in Education has never been offered at Hunter College. The average index in their various undergraduate major for those who earned a doctorate in Education is 2.79.

In most cases the average in the subject of the doctorate or a related major is higher than the average cumulative index which includes the major. In the physical and biological sciences there are, however, 12 cases where the subject index is lower, whereas in the social sciences there are only two, and none in the arts and humanities. In part this may account for the smaller differences between the cumulative indices and indices in the subject of the doctorate in the physical and biological sciences, 0.22 and 0.24 as compared with the social sciences and arts and humanities, 0.39 and 0.37 respectively. In the arts and humanities, one factor may be the number of individuals with foreign backgrounds who majored in the language of their native country, as the difference in indices for modern foreign language and literature is 0.50.

Distribution of Honors

The differences in grade indices reviewed above are reflected also in the data on honors, which are awarded, of course, on the basis of cumulative indices. Table V shows for each field the number and percent of those who received Junior and Senior Phi Beta Kappa, and graduation honors of summa cum laude, magna cum laude and cum laude. The physical sciences head the list with the arts and humanities second. Awards in Education are only slightly greater than those in the random sample, again pointing to the disparity seen in Figure 3. Mathematics, although included above in the physical sciences, is listed separately

TABLE III

COMPARISON OF THE AVERAGE UNDERGRADUATE CUMULATIVE INDICES AND THE AVERAGE UNDERGRADUATE INDICES IN THE SUBJECT OF THE DOCTORATE IN DIFFERENT FIELDS

Field	Number of Individuals	Average Cumulative Index	Number of Individuals	Average Index in Doctoral Subject	Difference between indices
Physical Sciences	37	3.26	37	3.48	.22
Arts and Humanities	74	3.17	73	3.54	.37
Biological Sciences	53	3.11	52	3.55	.24
Psychology	33	3.07	21	3.51	.44
Social Sciences	33	3.06	30	3.45	.39
Education	64	2.71	52	2.99	.28
Non-Education	230	3.18	213	3.47	.29
Total	294	3.04	265	3.37	.33

FIGURE 3

DISTRIBUTION OF THE CUMULATIVE INDICES OF STUDENTS WHO LATER EARNED THE DOCTORATE IN EDUCATION, OF STUDENTS WHO EARNED THE DOCTORATE IN SUBJECTS OTHER THAN EDUCATION AND OF STUDENTS IN THE RANDOM SAMPLE

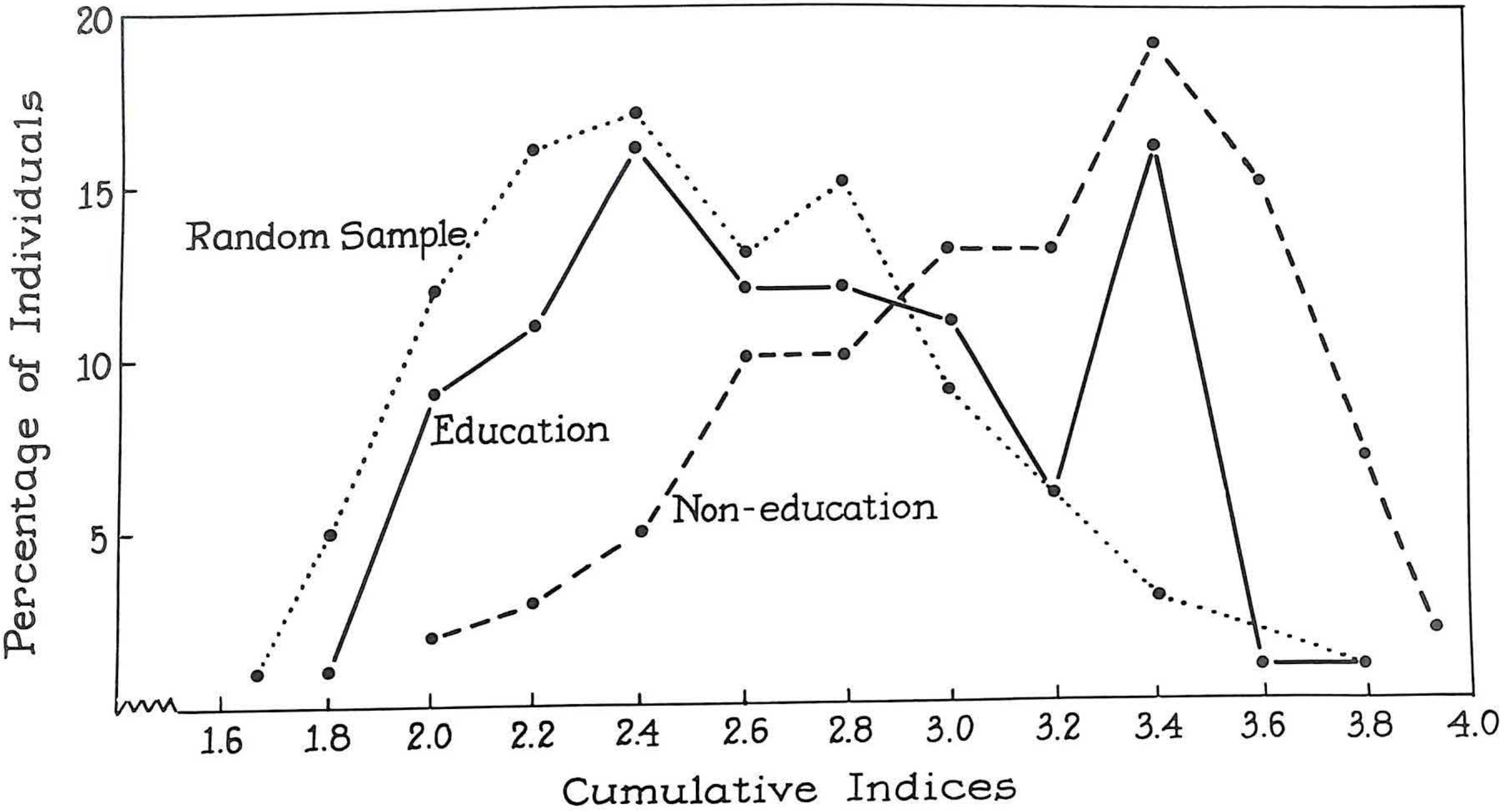


TABLE IV

COMPARISON OF THE AVERAGE CUMULATIVE INDICES FOR EACH SUBJECT IN WHICH THERE WERE THREE OR MORE STUDENTS WHO EARNED THE DOCTORATE

Subject	Number in Group	Average Cumulative Index
Mathematics	8	3.61
Classical Languages	7	3.47
Anthropology	6	3.28
Law and Jurisprudence	4	3.25
Physics	8	3.23
Biochemistry	8	3.18
English and American Literature	18	3.16
History	10	3.15
Modern Foreign Languages and Literature	29	3.11
Chemistry	17	3.09
Biology	37	3.02
Psychology	33	3.02
Economics	7	3.00
Philosophy	5	3.00
Physiology	6	2.95
Geology	3	2.83
Sociology	6	2.81

Analysis of variance gives a value for F of 1.91 and for P of less than .05.

TABLE V

DISTRIBUTION OF UNDERGRADUATE HONORS AMONG THE STUDENTS IN THE DIFFERENT FIELDS IN WHICH DOCTORATES WERE EARNED

Field	Number in Group	Phi Beta Kappa				Graduation Honors				
		Junior	Senior	Total	% of group	Summa cum laude	Magna cum laude	Cum laude	Total	% of group
Physical Sciences	37	13	8	21	57%	3	7	10	20	54%
Arts and Humanities	74	15	16	31	42%	3	11	24	38	51%
Biological Sciences	53	9	11	20	38%	1	4	17	22	42%
Psychology	33	2	10	12	36%	2	2	7	11	33%
Social Sciences	33	5	6	11	33%	2	1	8	11	33%
Education	64	1	4	5	8%	0	0	11	11	17%
Total	294	45	55	100	34%	11	25	77	113	39%
Non-Education	230	44	51	95	41%	11	25	66	102	44%
Sample	294	5	20	25	9%	1	4	25	30	10%
Mathematics	8	7	0	7	88%	3	4	0	7	88%

because of the outstanding record of seven high awards in each category conferred within a group of only eight undergraduates.

Male Versus Female Students

Of the 13 male students included in this study, 12 admitted in 1946 or thereafter graduated in 1949 or 1950, and so had obtained the doctorate by 1956 when this study ends, in a maximum of six or seven years. Their cumulative indices ranged from 3.3 to 3.7, with an average of 3.4. Only three women of those graduating in 1949 or 1950 had earned the doctorate by 1956, with individual indices of 2.8, 2.9 and 3.9 and an average of 3.2. During this period there were 357 male graduates as compared with a total of 2,358 female graduates, a ratio of one to 6.61. Thus the ratio of graduates earning the doctorate within six or seven years was approximately 26.4 men to one woman. This suggests that qualified women, due to marriage and their different role in society, tend to delay the pursuit of a doctorate or do not undertake it at all.

Discussion

The first objective of this study of the records of undergraduates from a single college who later earned a doctoral degree was to determine the quality of these students as compared with a random sample of the student body. The latter sample was composed of the same number of individuals for each class, selected randomly as described above, as the classes had contributed to the doctoral group. The results show that the average of the undergraduate records of the doctoral group is considerably higher than the average of the random sample with a difference (0.48) that is definitely significant ($P<0.001$). The distribution of the grades of the doctoral group is represented by a curve skewed to the left while the curve of the random sample is skewed to the right (Figure 1).

A comparison of the indices of the first term of the two groups presents a similar picture (Figure 2). In the latter chart, the tails in each case extend farther to the left than in Figure 1 because lower averages may occur in the first term than are permissible for graduation, but the means of the two curves differ by a decidedly significant amount, the future doctors earning an average index higher by 0.51 ($P< 0.001$). The correlation of the first term indices of the doctoral group with their cumulative indices is very high (0.86) and appreciably greater than the high correlation value for these two sets of indices for the random sample (0.71), indicating a greater degree of consistency in the records of the superior group.

The second objective was to determine whether the undergraduate records of students who later obtained the doctorate showed any significant differences as related to the particular disciplines in which they obtained the doctorate. When the records were separated into the six fields designated in the report of the Office of Scientific Personnel, no significant difference was noted except in the field of Education. When the individual disciplines were arranged separately in order of descending indices, significant differences were apparent between the extremes of the list. The groupings of subjects by the method of the Office of Scientific Personnel had obscured these differences, e.g., both Mathematics and Geology are included in the field of the Physical Sciences, but the average of students in Mathematics differs appreciably from that of students in Geology.

The average of the cumulative indices of students who later obtained a doctorate in Education differs from that of the non-education group by an amount which is definitely significant (average indices less by 0.47, with $P< 0.001$.) The distribution of the indices for the education group is represented by a bimodal curve (Figure 3) with one mode, of a smaller proportion, coinciding with the mode of the non-education group, and a second mode, composed of the majority, coinciding with the mode of the random sample. Webb (4) has found a similar difference among candidates for the Master's degree at Emory University. He reports that non-education students had better undergraduate records and made higher scores on all of a series of tests than did Education students.

Finally it was deemed advisable to consider these results for any criteria which might be applicable to the selection of students who possess the potential for advanced graduate work. Undergraduate records are, of necessity, the only indicator here. Relative attainments within the Graduate School have not been considered and the eventual awarding of the Doctor's degree has been accepted as a single criterion of success in graduate work. Some studies have approached the problem of undergraduate-graduate correlation from the point of view of ratings within the Graduate School. Webb (5) studied graduate students at Emory University, coming from a great variety of undergraduate institutions. He examined the relative merits of 1) undergraduate transcripts, 2) the Cooperative English Test, 3) the Cooperative General Culture Test, and 4) evaluation by the faculty of the Graduate School. The high value of the undergraduate transcripts is evident from his studies, even though his analysis dealt with undergraduates coming from numerous colleges of unlike standards. The undergraduate records examined for this data have been earned in a single institution where faculty appraisal and ratings are probably more homogeneous.

In this survey there is evidence that the cumulative index and even the first term index may be highly useful as a predictor of degree attainment in Graduate School. Approximately all students who have an index of 3.5 or more in the first term have the potential for high academic achievement, and those with a first term average of 3.0 to 3.4 include a large percentage who are qualified for graduate study. It would appear therefore to be highly advantageous to encourage students in these categories to consider graduate study and to adopt a curriculum in college that will provide better than average stimuli.

Conclusions

The undergraduate records of 294 students from one college who subsequently earned the doctorate were significantly better than those of a comparable random sample from the same classes, both for the first term and for the cumulative average of all their college work.

The correlation between first term grades and the cumulative average is high for both groups, with a coefficient of 0.86 for the doctors and 0.71 for the random sample. There is also significant correlation between high school and college averages for the future doctors with a coefficient of 0.66.

A few students with very poor averages earned the doctorate, including four who had been dropped for poor scholarship and were later reinstated.

Superior students are in almost all cases superior in their first term. Seventy-eight percent of the doctoral group with a first term index of 3.5 completed their college work with a 3.5 index and 92% with a first term index of 3.0 or above completed their work with an index equally high.

There is no significant difference between the college averages of the different fields as defined by the Office of Scientific Personnel, except for Education, in which the average was considerably below that of the other groups with a highly significant difference. The distribution of the undergraduate grades of the Education group fall into a bimodal curve with two modes, one for a larger group the same as that of the random sample and a second mode for a smaller group the same as that of the other doctors.

The grade averages when grouped according to subjects (rather than fields), with the omission of Education, range from a high of 3.61 for Mathematics to a low of 2.81. Analysis of variance indicates that these differences are significant.

The averages of the grades earned in the subject of the future doctorate or a related major are somewhat higher than those of the general average, a tendency which is more marked in the social sciences and the arts and humanities than in the physical sciences and the biological sciences.

The distribution of honors parallels the differences in averages in the different groups, with honors in Education fewer and comparable to the those in the random sample.

For graduates during 1949-50 when both male and female students received degrees, the number of male students obtaining the doctorate by 1956 was far in excess of the number of female students with a ratio of 1 to 26.4 when pro-rated for numbers of male and female graduates.

FOOTNOTES

1. Skewness of doctors as measured for β_1 was found to be 0.125 and skewness of sample was 0.216. For N = 300 both are beyond the 0.02 point. The skewness of each is, therefore, significant.
2. There is a factor of dependence involved in these correlations since the first term grades are included in the cumulative index, but the values are sufficiently high not to abrogate the correlation.
3. Data did not indicate exact degree earned, i.e., Ph.D. or Ed.D.

REFERENCES

1. Croxton, F. E. and Cowden, D. J., Applied General Statistics, Revised Edition, (Englewood Cliffs, New Jersey: Prentice Hall, Inc., 1955).
2. Dixon, W. J. and Massey, F. J., Introduction to Statistical Analysis, (Eugene, Oregon: Private Printing, 1949), pp. 1-5, reprinting Table XXXIII from Ronald A. Fisher and Frank Yates, Statistical Tables for Biological, Agricultural, and Medical Research, (Edinburgh: Oliver and Boyd, 1943).
3. Office of Scientific Personnel, National Academy of Sciences-National Research Council, "Production of U. S. Scientists: Trends in the Number of Doctorates Granted in the Major Fields of Science and in the Humanities," Science, CXXVII (28 March, 1958), pp. 682-685.
4. Webb, S. C., "Predictors of Achievement in Graduate School," Journal of Applied Psychology, XXXV, (August, 1951), pp. 265-271.
5. Webb, S. C., "Differential Prediction of Success in Graduate School," Journal of Educational Research, L, (September, 1956), pp. 45-54.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 1, September 1960)

SYLLOGISTIC REASONING

GEORGE N. BEAUMARIAGE, Jr.
Sacramento State College
Sacramento, California

REASONING HAS been considered as one of the factors of intelligence by many psychologists. Reasoning in most intelligence tests includes the induction of a rule from several instances and the drawing of conclusions from syllogisms. At times the induction of a rule may appear to be a type of matching and recognition of absurdities. Certain types of reasoning ability are directed toward, and are measured in achievement tests, as for example, understanding. This understanding factor seems to be in reality a type of reasoning ability. Some concept formation tests really measure the ability to choose the answer to a problem.

In the past, measurements of intelligence were more preoccupied with the ability necessary to apply formal logic. The English still seem to demand an ability to handle formal logic as a measure of "intelligence." Actually, in one of our more modern tests of mental maturity and ability, the California Mental Maturity Test, formal logic is found in only one-seventh of the test.

The problem of language has great importance when one considers logic as a measure of reasoning ability. If the test problem measuring logic involves the ability to understand written language, the logic might be obscured by the language ability of the testee.

For the purposes of this paper, syllogistic reasoning will be that ability necessary to logically determine the correct conclusion from a group of facts presented in the form of a syllogism.

The particular form of logic used will be syllogistic reasoning, and will, in effect, be a measure of the subject's ability to do deductive reasoning. Inductive reasoning will not be considered.

To determine the syllogistic reasoning abilities of the subjects, the following syllogisms were used: Those that are found in Part 4 of the California Short Form Test of Mental Maturity, Elementary 1950 S Form; and Sir Cyril Burt's Test No. 29, Graded Reasoning Test, Short List (2). Some changes in the language of the Burt list have been made to conform with American usage since this test was originally designed for English usage. A copy of the Graded Reasoning Tests (Short List) is included in the Appendix.

Some fifty-three pupils were given the California Mental Maturity Test and the Graded Reasoning Test, Short List, by Sir Cyril Burt. The Inference Test from the California Mental Maturity Test was then compared to the Graded Reasoning Test.

The Procedure

The California Mental Maturity Test is a group test. The writer administered this test to four groups of children composed of 4th, 5th, and 6th graders.

Approximately one and one-half hours were spent in administering each of these tests. The results from the testing of fifteen 4th graders are summarized in Table I, of fourteen 5th graders in Table II, and of twenty-four 6th graders in Table III. It is to be noted that these tables record the mental age score in months. The test itself is made up of spatial relationships, two parts of logic reasoning, two parts of numerical reasoning, and two parts of verbal conceptions according to the titles given. For the purposes of this paper the results from Test Four, Inferences, are recorded. This Test consists of fifteen situations that require logical application of deductive reasoning for solution. The test results, in terms of mental age in months, for this test are found in Column 5 of the tables.

Sir Cyril Burt's Test 29, Graded Reasoning Test, Short List, consists of some seventeen items. Starting at a base level of six and one-half years equal to 0 and counting each test correctly solved and explained as one-half year, the test goes to 15 years. Theoretically, a child with an IQ of 100, or average, will correctly answer all of the questions listed beneath his age. Each test is equivalent of one-half year with 0 falling at six and one-half.

This test requires individual examination. The child is required to read the question and may be helped with any and all language difficulties, until

TABLE I

FOURTH GRADE
Mental Age in Months

1	2	3	4	5	6	7	8	9	10	11
Child	CA	Spatial Relations	Logical Reasoning		Numerical Reasoning	Verbal Concepts	Language	Non-Language	Total Mental	B Syl.
			Syl.	Inf.						
1	121	119	132	94	133	120	110	126	116	96
2	131	134	176	168	120	118	126	134	132	156
3	121	134	107	131	133	108	114	123	119	138
4	121	123	155	179	141	134	141	126	135	156
5	123	102	155	101	95	0	0	107	92	102
6	119	158	155	108	132	126	117	150	132	102
7	117	84	132	109	120	108	103	96	102	114
8	133	84	97	180	120	123	127	90	121	162
9	120	120	131	92	130	2	10	122	90	102
10	132	132	170	160	122	118	124	119	121	144
11	118	115	110	130	132	122	118	120	119	138
12	119	131	150	160	138	116	128	130	130	162
13	123	87	148	110	100	130	105	103	104	102
14	122	96	146	120	128	108	110	108	110	108
15	134	86	100	115	115	120	108	110	109	96

TABLE II

FIFTH GRADE
Mental Age in Months

1	2	3	4	5	6	7	8	9	10	11
			Logical Reasoning							
Child	CA	Spatial Relations	Syl.	Inf.	Numerical Reasoning	Verbal Concepts	Language	Non-Language	Total Mental	B Syl.
1	131	123	114	114	96	128	113	118	117	117
2	138	108	175	108	120	125	118	112	116	102
3	136	177	156	122	134	118	101	158	120	114
4	133	195	174	188	102	135	135	162	148	159
5	127	147	156	130	159	135	134	156	142	126
6	125	127	176	130	134	143	136	142	138	132
7	131	168	176	180	166	169	168	167	168	144
8	129	168	174	156	166	142	144	168	150	150
9	138	196	197	174	182	167	170	192	179	156
10	137	178	174	169	166	155	160	162	160	162
11	132	168	155	197	166	166	169	165	168	168
12	138	143	156	122	146	142	140	139	141	108
13	134	155	174	177	155	159	159	158	158	162
14	135	177	175	197	177	166	173	174	175	176

TABLE III

SIXTH GRADE
Mental Age in Months

1	2	3	4	5	6	7	8	9	10	11
Child	CA	Spatial Relations	Logical Reasoning Syl.	Logical Reasoning Inf.	Numerical Reasoning	Verbal Concepts	Language	Non-Language	Total Mental	B Syl.
1	143	176	154	178	147	116	132	160	140	168
2	135	148	156	121	139	152	141	148	144	114
3	157	169	174	177	159	151	156	167	158	174
4	147	176	174	169	158	152	158	162	158	162
5	145	195	174	168	140	146	146	174	153	132
6	147	105	175	186	180	149	165	123	149	176
7	144	107	174	114	159	144	143	114	133	103
8	136	186	156	157	184	146	153	181	165	162
9	145	154	106	114	148	122	116	145	128	102
10	138	140	132	132	172	148	148	139	147	126
11	151	186	175	121	141	112	115	194	136	126
12	140	104	155	123	173	131	138	116	130	114
13	168	186	174	157	160	156	156	164	157	144
14	152	96	177	102	125	132	118	108	114	96
15	145	185	155	143	167	157	155	166	157	150
16	146	98	174	132	156	137	143	108	128	132
17	146	128	174	120	126	160	137	138	138	114
18	143	160	156	144	134	144	144	140	156	138
19	156	123	156	131	158	139	140	132	138	132
20	149	196	174	144	114	148	134	175	148	132
21	156	140	174	108	113	151	133	142	138	114
22	143	124	114	156	134	150	150	114	138	144
23	165	156	132	127	127	128	144	144	138	126
24	150	134	174	168	140	109	128	132	131	162

TABLE IV

FOURTH-GRADE TEST SCORE COMPARISONS

1 C. A.	2 Inference	3 Total Mental	4 Rank Order Total Mental	5 B Syllogisms	6 Rank Order Syllogisms
133	180	121	4	162	1
121	179	135	1	156	2
131	168	132	2	156	2
132	160	121	4	144	3
119	160	130	3	162	1
121	131	119	5	138	4
118	130	119	5	138	4
122	120	110	7	108	6
135	115	109	8	96	8
123	110	104	9	102	7
117	109	102	10	114	5
119	108	132	2	102	7
123	101	92	11	102	7
121	94	116	6	96	8
120	92	90	12	102	7

TABLE V

FIFTH-GRADE TEST SCORE COMPARISONS

1 C. A.	2 Inference	3 Total Mental	4 Rank Order Total Mental	5 B Syllogisms	6 Rank Order Syllogisms
135	197	175	2	176	1
132	197	162	3	168	2
133	188	148	7	159	4
131	180	168	3	144	7
134	177	158	5	162	3
138	174	179	1	156	5
137	169	160	4	162	3
129	156	150	6	150	6
127	130	142	8	126	9
125	130	138	10	132	3
138	122	141	9	108	12
136	122	120	11	114	11
131	114	117	12	117	10
138	108	116	13	102	15

TABLE VI

SIXTH-GRADE TEST SCORE COMPARISONS

1	2	3	4	5	6
C. A.	Inference	Total Mental	Rank Order Total Mental	B Syllogisms	Rank Order Syllogisms
175	186	149	6	176	1
143	178	140	9	168	3
157	177	158	2	174	2
147	169	158	2	162	4
145	168	153	5	132	8
150	168	131	13	162	4
136	157	165	1	162	4
168	157	157	3	144	6
143	156	138	10	144	6
149	144	148	7	132	8
143	144	156	4	138	7
145	143	157	3	150	5
138	132	147	8	126	9
165	132	138	10	126	9
146	132	128	15	132	8
156	131	138	10	132	8
140	123	130	14	114	10
135	121	144	8	114	10
151	121	136	11	126	9
146	120	138	10	114	10
145	114	128	15	102	11
144	114	133	12	102	11
156	108	138	10	114	10
152	102	114	16	96	12

the tester is assured of his success with the language of the question. The subject may read silently if he desires and the tester finds that he can successfully handle the material.

After the child understands the task, he is allowed to think it through. Then he must answer the problem. After he has answered the problem, he is given a score of one. He must then give his reasons for the answer. If his reasoning is faulty, then one-fourth score is deducted for each faulty reason. If the child gives an incorrect answer to the problem, he is asked to try again. After four unsuccessful attempts, he continues to the next problem. After three consecutive problems are missed four times each, the test is finished. A correct answer may be cancelled as far as scoring is concerned by incorrect reasoning. Motivation, approval, and other practices conducive to good rapport between the tester and subject should be practiced.

The writer had each of these seventeen items on a card and gave the cards to the child for him to read. It was necessary to change some words from the original test so that they would more nearly resemble those in common usage in this country. For instance, the word "cleverer" was changed to "smarter", because the children had difficulty with "cleverer." The word "blouse" was changed to "shirt," "walking stick" to "cane," and "cricket" to "ball." Locations such as France, the Scottish Hills, and sea coast were changed to Hawaii, Sierras, and Beach. Some other words that could have given difficulty were not changed because no difficulty was actually incurred. The results of this test are found in Column II in Tables I, II, and III.

According to Sir Cyril Burt, the efficient test for measurement of mental abilities is one that requires logical inference for solution. When the testing situation gives all of the facts required for the solution, thus eliminating special knowledge, empirically that test is the best measure of intelligence available.

Conclusions and Implications

Some fifty-three children were given both the California Mental Maturity Test and the Graded Reasoning Test (Short List). Fifteen fourth graders were tested, fourteen fifth graders, and twenty-four sixth graders.

The test results are found in Tables I, II, and III. Tables IV, V, and VI give some of the scores most appropriate to this study. It is to be noted that these tables are in rank order in terms of months of mental age and include chronological age.

For the fourth-grade scores, the Spearman rank difference correlation, rho, between the syllogism test and the inference test is 0.88.

For the fifth-grade scores, rho is 0.86 between the total and inference scores, and 0.91 between the syllogism and inference scores.

The rank order of the scores for the syllogisms agrees very well with the rank order of the inference scores from the Mental Maturity Test. However, the total mental score from the Mental Maturity Test does not agree as well. Chronological age seems to have even less correlation.

Since the Mental Maturity Test purports to test more than logical reasoning ability, then perhaps these factors do not necessarily correlate with reasoning abilities.

Little significant difference was found in the ability to handle syllogisms between children of grades four and six. The relationship between chronological age and ability to think logically seems to be determined by some sort of intelligence, native or otherwise. Actually, those children who, in general, had the highest score in terms of mental age did the best on the syllogisms.

Language had an effect on ability to handle the syllogisms. Many times the children asked questions concerning pronunciation and meaning of words. Actually, there seemed to be an unwillingness to tackle any unfamiliar word for fear it would be misleading in the solution of the problem. There seemed to be a general attempt to hunt for hidden meanings, and when the children found out they could ask questions they tended to ask more than they actually needed to ask. This slowed down the process and thus much time was required.

When the children gave wrong answers or wrong reasons for their answers, they then began to guess more than when they had success. Guessing seemed to increase with failure. The guessing did not exactly reflect illogical reasoning, it more nearly reflected no reasoning at all. When cautioned, the children seemed to use better logic than when allowed to proceed without any restraint.

A second test equivalent to the Burt test but to be administered as a group test requiring written answers might disclose whether the children really needed all the information they requested. It is felt that the security the children got from the ability to ask questions materially aided them and resulted in slightly higher scores.

In general, there seems little doubt that the syllogism will give an indication of the mental ability of the subject. However, there is another factor (or factors) here—that the writer has too few data to identify correctly.

REFERENCES

1. Anastasi, Anne, and J. P. Foley, Jr. Differential Psychology (New York: Macmillan Co., 1949).
2. Burt, Sir Cyril. Scholastic and Mental Tests (London: Oxford University Press).
3. Cronbach, L. J. Essentials of Psychological Testing (New York: Harper and Bros., 1949).
4. Hilgard, Ernest R. Theories of Learning (New York: Appleton-Century-Crofts, Inc., 1948).
5. Maier, Norman R. F. "Reasoning and Learning," Psychology Review, XXXVIII (1931), pp. 332-46.
6. McGeoch, John A. The Psychology of Human Learning (New York: Longman's Green and Co., 1942).
7. Ross, C. C. Measurement of Today's Schools (New York: Prentice-Hall, 1947).
8. Thouless, R. H. How to Think Straight (New York: Simon and Schuster, 1939).
9. Woodworth, Robert S. Contemporary Schools of Psychology (New York: Ronald Press, 1948).
10. Woodworth, Robert S., and S. B. Sells, "An Atmosphere Effect in Formal Syllogistic Reasoning," Journal of Experimental Psychology, XVIII (1935), pp. 451-60.

APPENDIX

GRADED REASONING TESTS
(Short List)

1. Tom runs faster than Jim;
Jack runs slower than Jim.
Which is the slowest of the three?

7 Years

2. Kate is cleverer than May;
May is cleverer than Jane.
Who is the cleverest—Jane, Kate, or May?

3. I have bought the following Christmas presents: a pipe, a blouse, some music, a box of cigarettes, a bracelet, a toy engine, a bat, a book, a doll, a walking-stick, and an umbrella.
My brother is eighteen; he does not smoke, nor play cricket, nor play the piano.
I want to give the walking-stick to my father, and the umbrella to my mother.
Which of the above shall I give to my brother?

8 Years

4. I don't like sea voyages;
And I don't like the seaside.
I must spend Easter either in France, or among the Scottish Hills, or on the South Coast.
Which shall it be?

5. The person who stole Brown's purse was neither dark, nor tall, nor clean-shaven.
The only persons in the room at the time were—
1) Jones, who is short, dark, and clean-shaven;
2) Smith, who is fair, short, and bearded;
3) Grant, who is dark, tall, but not clean-shaven.
Who stole Brown's purse?

9 Years

6. Three boys are sitting in a row;
Harry is to the left of Willie;
George is to the left of Harry.
Which boy is in the middle?

7. In cold, damp climates, root crops, like potatoes and turnips, grow best;
In temperate climates, there are abundant pastures, and oats and barley flourish;
In sub-tropical climates, wheat, olives, and vines flourish.
The ancient Greeks lived largely on bread, with oil instead of butter; they had wine to drink, and raisins for fruit.
Which climate do you think they had?

10 Years

8. There are four roads here:
I have come from the south and want to go to Melton;
The road to the right leads somewhere else;
Straight ahead it leads only to a farm.
In which direction is Melton—North, South, East, or West?

9. The doctor thinks Violet has caught some illness.
If she has a rash, it is probably chicken pox, measles, or scarlet fever;
If she has been ailing with a cold or cough, she may develop whooping cough, measles or mumps.
She has been sneezing and coughing for some days, and now spots are appearing on her face and arms.
What do you think is the matter with Violet?

11 Years

10. Where the climate is hot, gum trees and rubber will grow;
Heather and grass will grow only where it is cold;
Heather and rubber require plenty of moisture;
Grass and gum trees will grow only in fairly dry regions.
Near the Amazon River it is very hot and very damp.
Which of the above grows there?

11. Father has just come home in a brand new overcoat; there is clay on his boots and flour on his hat.
The only places he can have been to are Northgate, Southgate, Westgate, or the City; and he has not had time to go to more than one of these.
There is no clay anywhere in the streets except where the pavement is up for repair;
There are tailor's shops only in Southgate, Westgate, and the City;
There are flour mills only in Northgate, Westgate, and the City;
I know the roads are not being repaired in the City, though they may be in other places.
Where has father been?

12 Years

12. I started from the church and walked 100 yards;
I turned to the right and walked 50 yards;
I turned to the right again and walked 100 yards.
How far am I from the church?

13. Field mice devour the honey stored by the bumblebees; the honey which they store is the chief food of the bumblebees.
Near towns, there are far more cats than in the open country.
Cats kill all kinds of mice.
Where, then, do you think there are the most bumblebees—in the neighborhood of towns or in the open country?

13 Years

14. A pound of meat should roast for half an hour;
Two pounds of meat should roast for three-quarters of an hour;
Three pounds of meat should roast for one hour;
Eight pounds of meat should roast for two hours and a half.
From this can you discover a simple rule by which you can tell from the weight of a joint for how long it should roast?

15. What conclusions can you draw from the following facts?
Iron nails will not float in a pool;
A cup of pure gold dust weighs nearly twenty times as much as a cup of water of the same size;
If you drop a silver sixpence or a copper coin into a puddle, it will sink to the bottom;
A cubic inch (about a tablespoonful) of water weighs less than half an ounce; a cubic inch of brass weighs over two ounces;
A leaden weight will drop to the bottom of the ocean.
Sum up all these observations in one short statement of the following form:
"Most __________ are __________ ______ ________."

14 Years

16. John said: "I heard my clock strike yesterday, ten minutes before the first gun fired. I did not count the strokes, but I am sure it struck more than once, and I think it struck an odd number."
John was out all morning from the earliest hours; and his clock stopped at five to five the same afternoon.
When do you think the first gun fired?

17. Captain Watts and his son James have been found shot—the father in the chest and the son in the back. Both clearly died instantaneously.
A gun fired close to the person—as, for example, when a man shoots himself—will blacken and even burn the skin or clothes; fired from a greater distance, it will leave no such mark.
The two bodies were found near the middle of a large hall used as a rifle range. Its floor is covered with damp sand, which shows every footprint distinctly. Inside the room there are two pairs of footprints only. A third man standing just outside the door or window could aim at any part of the room, but the pavement outside would show no footmarks.
Under Captain Watts' body was found a gun; no such weapon was found near James.
In each case the coat, where the bullet entered, was blackened with gunpowder, and the cloth a little singed.
Captain Watts was devoted to his son, and would have died sooner than harm him purposely; hence it is impossible to suppose that he killed him deliberately, even in self-defense. But some think that James secretly disliked his father, and hoped to inherit his fortune at his death.
Was Captain Watts' death due to murder, accident, or suicide?
Was James' death due to murder, accident, or suicide?

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 1, September 1960)

ANALYSIS OF ADOLESCENTS' RESPONSES TO THE F SCALE

JACK R. FRYMIER
Auburn University

The Problem

WILL DIFFERENCES in cultural background affect the way adolescents respond to the F scale, a measure of anti-democratic (authoritarian) potential? If there are differences, will they be apparent in total F scale scores? Will differences also be apparent in individual items of the measure? If there are differences, what do they mean?

Christie and Garcia[1]* compared F scale responses of a group of California university students with a college population from a southwestern city. They reported that the scores from the California group were significantly lower and also that the California group accepted 18 of the 30 items to a lesser degree than the other group. Other studies by Prothro and Melikian[2] and Cohn and Carsch[3] indicate that persons living in or coming from an authoritarian culture tend to score higher on the F scale.

The Procedure

To answer the questions posed above, the author administered the F scale in conjunction with other tests[4] to 104 students in a large metropolitan high school in Michigan and a rural consolidated school in Alabama. For the purposes of this study, it was assumed that urban Michigan and rural Alabama constituted different cultural situations. Each student from the Northern urban area was matched with one from the Southern rural area on the bases of age, sex, and grade in school, making 52 pairs in all. The mean age of the two groups was 16. 38 years, and each group had 22 girls and 30 boys. These students had all lived in their particular locale for at least ten years and all were Caucasian. In response to preliminary questions about frequency of church attendance during a one-month period, the Northern group mean was 3. 74 and the Southern group mean was 4. 42 times. The Southern rural group also had more persons who indicated they watched television "very often" while the Northern urban group had a greater number who said they watched television "occasionally."

The responses to the F scale were tabulated in two ways. First, total scores were computed according to the directions of Adorno[5]. Next, agreement or disagreement[6] with the individual items was recorded.

These data were then manipulated statistically several ways. First, mean F scale scores and standard deviations were determined for each group, and the difference between these mean scores was then compared by use of the "t" test. Second, the number in the Southern rural and Northern urban groups who agreed and disagreed on each item was compared by computing chi square. Third, the percent of each group who agreed with each item was determined, and the various items listed in rank order of agreement for both groups. Fourth, these rank orders were compared by inspection. The results of these procedures follow.

The Results

The difference between the mean F scale scores for the Northern urban group (mean of 3. 95) and Southern rural group (mean of 5. 05) was statistically significant, as shown in Table I. This seems to indicate that there is some factor or factors which may be responsible for a considerable difference between these two groups on this particular measure, the Southern rural group apparently tending to exhibit more anti-democratic (authoritarian) potential than the Northern urban group.

Table II shows a comparison of the Southern rural and Northern urban groups' responses to individual items[7] of the F scale by use of chi square. Since the difference between the groups is statistically significant for 19 of the 27 items, this would seem to indicate that there is considerable difference in the way in which the members of these two cultural groups responded to these particular items of the F scale.

Table III shows the relative rank order for 22 of the 27 items of the F scale for both Northern and Southern groups. Since this ranking was not identical (there was not a perfect correlation for all

---

*All footnotes will be found at end of article.

TABLE I

COMPARISON OF MEAN SCORES OF SOUTHERN RURAL AND NORTHERN URBAN GROUPS ON F SCALE

Group	Mean	S. D.	"t"
Northern Urban	3. 95	. 620	8. 66*
Southern Rural	5. 05	. 667	

*Difference is significant at the . 001 level of confidence.

TABLE II

COMPARISON OF AGREEMENT AND DISAGREEMENT ON INDIVIDUAL ITEMS OF THE F SCALE FOR SOUTHERN RURAL AND NORTHERN URBAN GROUPS

Item Number[9]	Chi Square
1	. 00
2	7. 08*
3	. 98
4	7. 26*
5	. 34
6	10. 80*
7	. 96
8	10. 04*
9	10. 86*
10	5. 56*
11	12. 48*
12	3. 83
13	5. 63*
14	8. 09*
15	6. 22*
16	6. 19*
17	6. 71*
18	16. 12*
19	1. 46
20	. 96
21	13. 28*
22	9. 96*
23	22. 10*
24	15. 46*
25	3. 02
26	7. 96*
27	17. 71*

*Differences are significant at the . 05 level of confidence.

TABLE III

APPROXIMATE RANK ORDER OF F SCALE ITEMS FOR SOUTHERN RURAL AND NORTHERN URBAN GROUPS ACCORDING TO PROPORTION OF GROUP WHO AGREED WITH EACH ITEM

Item Number[9]	Item	North. Urban Agreement %	South. Rural Agreement %
1	Obedience and respect for authority are the most important virtues children should learn.	98	98
12	Young people sometimes get rebellious ideas, but as they grow up they ought to get over them and settle down.	88	98
2	No weakness or difficulty can hold us back if we have enough will power.	85	98
17	Someday it will probably be shown that astrology can explain a lot of things.	73	92
16	There is hardly anything lower than a person who does not feel a great love, gratitude and respect for his parents.	71	90
3	Science has its place, but there are many things that can never possibly be understood by the human mind.	77	85
8	What youth needs most is strict discipline, rugged determination, and the will to work and fight for family and country.	62	88
14	Sex crimes, such as rape and attacks on children, deserve more than mere imprisonment; such criminals ought to be publicly whipped, or worse.	60	85
25	No sane, normal, decent person could ever think of hurting a close friend or relative.	63	79
21	The wild sex life of the ancient Greeks and Romans was tame compared to the goings-on in this country, even in places where people might least suspect it.	54	86
22	If people would talk less and work more, everybody would be better off.	54	83
4	Human nature being what it is, there will always be war and conflict.	54	79
13	What this country needs most, more than laws and political programs, is a few tireless, courageous, devoted leaders in whom people can put their faith.	56	75

TABLE III (Continued)

Item Number[9]	Item	North. Urban Agreement %	South. Rural Agreement %
9	Nowadays when so many different kinds of people move around and mix together so much, a person has to protect himself especially carefully against catching an infection or disease from them.	50	71
6	When a person has a problem or a worry, it is best for him not to think about it, but to keep busy with more cheerful things.	42	75
7	A person who has bad manners, habits, and breeding can hardly expect to get along with decent people.	52	62
11	Some people are born with an urge to jump from high places.	35	69
20	Most of our social problems would be solved if we could somehow get rid of the immoral, crooked, and feeble-minded people.	44	54
10	An insult to our honor should always be punished.	35	58
15	People can be divided into two distinct classes, the weak and the strong.	23	52
24	The business man and the manufacturer are much more important to society than the artist and the professor.	15	52
26	Nobody ever learned anything really important except through suffering.	19	46

items, this requires some explanation. The author observed that although the proportion of the two groups who agreed with any particular item differed, in general both groups tended to accord approximately the same value to most of the various items as they related to the whole. That is, even though 75% of the Southern rural group agreed with item number 13 and only 56% of the Northern urban group agreed, if the percent of agreement for all of the various items were given a rank order for both groups, both the Southern rural and Northern urban groups accorded this particular item approximately the same relative position, viz., thirteenth. There were some variations in this relative rank order between the groups, but the general pattern seemed to be quite similar. For the purposes of this study, then, it was decided that if the relative position of a particular item did not vary more than four positions from that of the other group, this item was considered to have been accorded approximately the same value by both groups in the overall hierarchy of values which seems to be demonstrated by a complete ranking of all the items.

Table III, then, might be interpreted as giving some general indication of the extent to which these two groups were similar rather than different, as Tables I and II seem to indicate.

Differences greater than four positions in rank order placement were apparent for four of the 27 items. These are shown in Table IV. It should be noted that the relative difference in positions of item 19 for the two groups was not greater than four. There were only two items (numbers 5 and 19), however, on which the Southern rural group indicated less agreement than the Northern urban group, and this exception seemed important.

Discussion of Results

Now, what do all these data mean? Tables I and II seem to indicate that there are differences between the responses of these two cultural groups to the scale. Table III, on the other hand, seems to indicate a rather large area of general agreement, while certain specific differences are illustrated in Table IV.

Several conclusions, some conflicting, seem possible. The Southern rural group may be more authoritarian than the Northern urban group; or, the Southern rural group may be less intelligent or more prejudiced and the Northern urban group more intelligent and less prejudiced. Many persons[8] have pointed out the relationship between F scale scores and intelligence and prejudice. Perhaps all of these conclusions are warranted.

Next, there seem to be certain common values which are reflected in the relative amount of agreement with the individual items of the measure by both the Southern rural and the Northern urban groups, with some disagreement on a few other values. The author is not certain whether this second inference is justifiable or not, and if it is, whether it cancels out part or all of the first conclusion (i. e., regarding authoritarianism and/or intelligence and/or prejudice). It would seem to indicate, however, that the differences between these two groups may not be so great as one might assume from the statistical results, or else that they are of an entirely different nature than might otherwise be supposed.

A further observation might be that children from this particular Southern setting seem to be more reluctant to disagree than the children from the Northern area. Or, said another way, their pattern of responses seems to be one of "agreeing with" rather than "differing from" the test as a whole. There are at least three other important observations one might make from a study of Tables III and IV, however, and these are discussed in more detail below.

First, the general hierarchy of values evident upon inspecting the items in descending order of agreement seems to point out several general, though perhaps inconclusive, tendencies. For example, if one examines both ends of the rank order in Table III he gets the impression that these young people want and like mature assistance and direction, but they also have confidence in themselves and their own potential. Further, they seem to feel that people are important—that everybody counts.

Second, the fact that there is considerable difference in the extent to which both groups agreed with items 18, 23, and 27, (see Table IV), all of which are related to social intercourse in one way or another, seems to point up the fact that the members of this particular Southern rural group may be more distrustful, more afraid, and have less faith in other people than the persons from the Northern urban area. The statistical differences between the groups on these particular items (see Table II) are greater than the differences on any of the other items, too.

Also, items 5 and 19, which were the only two with which the rural group indicated less agreement than the urban group, seem to be related to religious tenents or beliefs. That the Southern rural group tended to give less value to these items than the Northern urban group may be an indication of some rebellion at religion on their part. If this is so, it would not be in keeping with the general pattern which seems to be apparent, i. e., the Southern rural group being more authoritarian. As such it would have to be explained some other way.

Third, an analysis of the responses to individual items according to sex indicates a general pattern of similarity within each of the two cultural groups, with certain exceptions. These exceptions are discussed below.

Northern urban girls agreed more than North-

TABLE IV

F SCALE ITEMS ON WHICH RURAL AND URBAN GROUPS DIFFERED

Item Number[9]	Item	North. Urban Agreement %	South. Rural Agreement %
18	Nowadays more and more people are prying into matters that should remain personal and private.	62	94
23	Most people don't realize how much our lives are controlled by plots being hatched in secret places.	42	86
27	Familiarity breeds contempt.	35	75
5	Every person should have complete faith in some supernatural power whose decisions he obeys without question.	62	56
19	Wars and social troubles may someday be ended by an earthquake or flood that will destroy the whole world.	44	35

ern urban boys on items 3, 5, 14, and 22—those concerned with the limited value of science, faith in the supernatural, punishing sex crimes, and working more and talking less. On the other hand, the boys from this Northern urban area tended to agree more than the girls from there on items 10, 11, 13, 16, and 23—those related to punishing insults, jumping from high places, need for leaders, love for parents, and control of our lives by plots.

Southern rural girls tended to indicate more agreement than boys from that area on items 13 and 16—need for leaders and love for parents. The boys from this particular rural area, however, agreed more on items 6, 7, 14, 15, and 20—keep busy and don't worry, need for good manners, punishing sex crimes, people are weak or strong, and getting rid of immoral people.

These differences can probably be accounted for, in part at least, by the experiences, desires, and hopes which these students projected into their responses. That the Alabama girls disagree more regarding punishment of "sex crimes such as rape" than their counterparts from the city or than the rural boys is one especially interesting example of this idea.

One final question seems to be in order: Is the F scale valid? Employment of the F scale in the conventional manner may point up statistical differences which may or may not be significant otherwise. If values are related to authoritarian potential, and such would seem to be a reasonable assumption, and if ranking items in order of agreement gives any indication of a hierarchy of values, it would appear that one may be able, in certain instances, to demonstrate more areas of agreement in responses to the F scale than disagreement; more similarities in values than differences. Granted, the procedure employed in this study was only a rough approximation of the extent to which these two groups were similar on this matter. It may be that more refined procedures might be developed or employed in future studies. This author feels compelled to express certain reservations about the validity of the F scale as a measure of anti-democratic or authoritarian potential at this point, however. It is admitted that these reservations are more vague than specific, but there seem to be some inconsistencies involved.

Suggested Research

Several interesting problems were raised during the course of this study, but the two which seemed most significant to this writer were as follows: 1) How would persons from a Northern rural area and from a Southern urban area respond to the various items of the F scale? 2) How would those persons who work with adolescents expect high-school students to agree with the various items of the F scale? Stated another way, would teachers or prospective teachers or social workers, for example, accurately perceive adolescents' values?

Summary and Conclusions

Two matched groups of 52 high-school students each from urban Michigan and rural Alabama were administered the F scale. The statistical differences between the two groups on the F scale mean scores and 19 of the individual item scores were significant. Both groups, however, tended to rank 23 of the 27 items in approximately the same relative position, as indicated by the proportion of each group who agreed with the particular item. The four items which were not generally agreed upon seemed to indicate that the Southern rural group expressed more fear and possibly distrust of other people than did the Northern urban group. Further, there also seemed to be some indication of a negative reaction to certain traditional religious ideas on the part of the Southern rural group.

The areas of agreement, however, seemed greater, and generally pointed up the idea that young people from both cultural situations possessed genuine concern for their fellow men coupled with considerable faith in their own ability to get things done. They seemed to want and to expect counsel and assistance, in return for which they were willing to accord a reasonable amount of love and respect for their elders. Expressed in one word, these adolescents seemed to possess that priceless human ingredient—hope.

FOOTNOTES

1. Richard Christie and J. Garcia, "Subcultural Variations in Authoritarian Personality," Journal of Abnormal and Social Psychology, XLVI (October 1951), pp. 457-69.
2. E. T. Prothro and L. Melikian, "The California Public Opinion Scale in an Authoritarian Culture," Public Opinion Quarterly, XVII (Fall 1953), pp. 353-62.
3. T. S. Cohn and H. Carsch, "Administration of the F Scale to a Sample of Germans," Journal of Abnormal and Social Psychology, XLIX (July 1954), p. 471.
4. These other tests were tests of hearing acuity and aural perceptions. See Jack R. Frymier The Relationship of Certain Behavioral Characteristics to Perception, unpublished Ed.D. dissertation, University of Florida, 1957.
5. T. W. Adorno, et al., The Authoritarian Personality (New York: Harper and Brothers, 1950), p. 265.
6. Any item which received a plus mark (+1 +2 +3) was considered as agreement, whereas any item which received a minus mark (-1 -2 -3) was considered as disagreement.

7. For the sake of brevity, items are listed only once. See Tables III and IV for a description of each item.
8. Edwin H. Titus and E. P. Hollander, "The California F Scale in Psychological Research: 1950-1955," Psychological Bulletin, LIV (January 1957), pp. 47-64.
9. Item numbers here refer to the author's numerical order as they were included in the version of the F scale which he used, rather than Adorno's item numbers as they are frequently referred to in the literature.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 1, September 1960)

COMPARISON OF ACADEMIC ACHIEVEMENT, APTITUDES, AND INTEREST PATTERNS OF TWO-YEAR TECHNICAL STUDENTS AND FOUR-YEAR DEGREE CANDIDATES IN ENGINEERING

LOUIS M. HERMAN and MARTIN L. ZEIGLER
The Pennsylvania State University

THE CURRENT emphasis on engineering as a career has helped produce a bumper crop of applicants for college training in this area. Since the available colleges and universities can handle only a portion of those applying for admission, the problem of proper selection and placement of potential students has become acute. The overall field of engineering is in need of both fully trained engineers prepared to handle complex problems and of technicians prepared to serve as engineering aides.

To meet the demands for this dual level of training, the Pennsylvania State University conducts a four-year engineering curriculum at its main campus and also a two-year technical program at its various centers. The study reported in this paper is an evaluation of the academic achievements, aptitudes, and interests of students within these four- and two-year curriculums. The two purposes of the study are 1) to gain information useful for proper selection and curriculum placement of students interested in engineering and 2) to determine the usefulness of a dual program in engineering.

Subjects and Procedure

Seven hundred and fifty-three freshmen enrolled in the four-year program at the College of Engineering and Architecture at the main campus in the fall of 1956, and 416 freshmen enrolled in the two-year technical program in engineering at the college centers during that same period comprised the two main curriculum groups studied. These numbers represent all students who completed their first semester. (In the remainder of the report, the four-year engineering program will be referred to as curriculum 1, and the two-year technical program as curriculum 2). Members of curriculum 1 and curriculum 2 were categorized into three subgroups called high achievers, low achievers, and withdrawals, according to the following criteria:

1. High achievers--those remaining in their curriculum beyond their first semester and maintaining a grade point average of two or better.[1]*
2. Low achievers--those remaining in their curriculum beyond their first semester, but falling below a grade point average of two.
3. Withdrawals--those who withdrew from their curriculum before their second semester began.

Aptitudes of the various subgroups were compared, using the Pennsylvania State University Academic Aptitude Examination (2). Subtests compared on this examination were the arithmetic and algebra sections and a four-part total consisting of a summation of scores on these two previous sections plus scores on the vocabulary and paragraph reading sections.

Interests of the various subgroups were compared with 24 occupations or groups of occupations on the Strong Vocational Interest Blank for Men.[2] Both the ability test and the interest test were administered when the students first enrolled at the university, before any course work was undertaken.

Results and Discussion

Achievements. The numerical and percentage breakdown of students in the two curriculum groups into the three subgroups is shown in Table I. The overall chi-square value of 22.67 for differences in distribution of students among the subgroups of the curriculums was significant at better than the .001 level. The percentage difference between curriculum 1 high achievers and curriculum 2 high achievers was not significant ($\underline{CR} = 1.23$, $\underline{p} < .20$). There were, though, a significantly greater percentage of low achievers in curriculum 2 than in curriculum 1 ($\underline{CR} = 2.39$,

---

*Footnotes will be found at the end of the article.

TABLE I

THE NUMBER AND PERCENTAGE OF HIGH ACHIEVERS, LOW ACHIEVERS, AND WITHDRAWALS IN CURRICULUM 1 AND IN CURRICULUM 2

Group	High Achievers		Low Achievers		Withdrawals	
	Number	Percent	Number	Percent	Number	Percent
Curriculum 1	394	52.3	210	27.9	149	19.8
Curriculum 2	233	56.0	144	34.6	39	9.4

$p < .02$), but a significantly greater percent of withdrawals in curriculum 1 than in curriculum 2 (CR = 4.24, $p < .001$). Apparently then, the percentage of clear-cut success in both groups is about equal, but students in curriculum 1 tend to drop out of the program to a greater degree than students in curriculum 2. The greater percentage of students remaining in curriculum 2 probably accounts for the greater percentage of low achievers in this group than in curriculum 1.

Aptitudes. Table II compares the mean scores obtained on the aptitude subtests by the various subgroups. It is seen that in every case curriculum 1 subgroups achieved higher mean scores than the corresponding curriculum 2 subgroups. Since there were nine comparisons, the probability of such an event occurring by chance is $(1/2)^9$ or less than .002. It is also seen that the lower mean score obtained by any curriculum 1 subgroup exceeds the highest mean score obtained by any curriculum 2 subgroup. Clearly then, curriculum 1 students manifest greater abilities as measured by these tests than do curriculum 2 students.

Comparisons of scores on the four-part total of the aptitude test were made within subgroups of curriculum 1 and within subgroups of curriculum 2. For curriculum 1, an analysis of varaiance showed an overall significant difference between subgroups ($F = 25.77$, $p < .001$). Subsequent critical ratio tests for significance of differences between any two subgroups showed that the high achievers earned a significantly higher test score than either the low achievers or the withdrawals ($p < .001$); differences between the low achievers and the withdrawals, however, were not significant.

Similar results were obtained within curriculum 2. An analysis of variance showed an overall significant difference between subgroups ($F = 52.41$, $p < .001$), and subsequent critical ratio tests revealed that high achievers exceeded at the .001 level of significance the test scores of both the low achievers and the withdrawals. Again, differences in test scores between the low achievers and the withdrawals were not significant.

It may be surmised from these findings that the differences in grades attained by the subgroups or their decision to withdraw may reflect differences in basic abilities. Within a curriculum group, the students withdrawing and the low achievers seem to have about equal abilities, but the high achievers are clearly superior to both. In a previous study by Guthrie (1), a correlation of .49 was found for four-year engineering students between scores on the four-part total of the ability test and grade-point averages. Although no correlational study was made on two-year technical students, it is quite probable that a similar relationship exists within that curriculum. It would seem likely then, that many of the curriculum 1 low achievers and withdrawals might achieve high grade-point averages within curriculum 2. On the other hand, it is probable that the majority of curriculum 2 high achievers would not achieve similarly high grade-point averages in curriculum 1.

Interests. The Strong Vocational Interest Test yields interest scores which can be expressed in letter grades of A, B+, B, B-, C+, C, and C-; respectively, these grades can be considered to range from indications of interests very close to the majority of people already engaged in a particular occupation to indications of interests quite dissimilar to the majority of people in that occupation. According to Strong's viewpoint that "when hiring a man for a position, it is recommended that he be considered to have passed the interest test for that occupation if he obtains an A or a B+ rating..."(3, pp. 3-4), there were only five areas out of the 24 examined in which any subgroup "passed" the interest test of an occupation.

These five areas were chemist, engineer, mathematics-science teacher, farmer and occupational group II (a composite of the first two areas). In both the chemist and the engineer categories, only the high and low achievers in curriculum 1 and the high achievers in curriculum 2 passed the interest tests. In the mathematics-science teacher category, these same three groups passed plus the withdrawals in curriculum 1. All subgroups passed the interest tests in the farmer category and in the group II composite. It seems then, that there is some tendency for achievement in the curriculums and interest in engineering and chemistry to go together, and to a somewhat lesser extent, for achievement and interest in teaching mathematics and science to go together. Where interests differentiate then, the distinction seems to be more on the basis of degree of academic success than on curriculums.

Table III shows the degree of similarity of interests of all subgroup comparisons. In the body of the table are shown the number of interest areas of the 24 measured in which two compared subgroups received identical median letter grades. The rows show the number of identical median letter grades and the columns under which the two numbers in a row are located indicate the two subgroups being compared. Thus, it is seen at the top of the table body, for example, that there were 20 out of 24 interest areas in which curriculum 1 high achievers and curriculum 1 low achievers received identical median letter grades and, at the bottom of the table, that there were nine interest areas out of 24 in which curriculum 1 low achievers and curriculum 2 low achievers received identical median letter grades.

In considering possible transfer of students between curriculums, the important findings revealed by the table are that curriculum 1 low achiev-

TABLE II

COMPARISON OF MEAN SCORES ON THE ABILITY TEST OF STUDENTS IN CURRICULUM 1 AND STUDENTS IN CURRICULUM 2

Subgroup and Test Section	Curriculum 1		Curriculum 2	
	Mean	S.D.	Mean	S.D.
High Achievers				
Arithmetic	24.99	4.03	19.50	4.25
Algebra	28.62	3.14	19.87	6.56
Four-Part Total	140.46	21.61	112.95	22.86
Low Achievers				
Arithmetic	22.14	4.17	15.68	4.14
Algebra	26.46	3.55	15.98	5.77
Four-Part Total	127.50	20.03	92.76	20.35
Withdrawals				
Arithmetic	22.09	4.80	16.86	5.18
Algebra	26.03	4.24	15.95	5.76
Four-Part Total	124.85	21.58	98.00	22.90

TABLE III

THE NUMBER OF INTEREST AREAS OF THE TWENTY-FOUR MEASURED IN WHICH THE VARIOUS PAIRED SUBGROUPS OBTAINED IDENTICAL MEDIAN LETTER GRADES

Curriculum 1			Curriculum 2		
High Achievers	Low Achievers	Withdrawals	High Achievers	Low Achievers	Withdrawals
20	20				
		18			18
		17		17	
				17	17
16			16		
	16		16		
		16	16		
			15		15
			14	14	
	12				12
11				11	
11					11
	11	11			
9		9			
	9			9	

ers are quite similar in interest to curriculum 1 high achievers, and to a somewhat lesser extent, to curriculum 2 high achievers. On the basis of interest then, curriculum 1 low achievers would probably adapt well enough within curriculum 2.

The curriculum 1 withdrawals are most like the curriculum 2 low achievers and curriculum 2 withdrawals. It is possible then, that if these latter two groups are disinterested in curriculum 2 (as evidenced by withdrawal or substandard achievement), the curriculum 1 withdrawals might also be disinterested. It should be noted though, that these curriculum 1 withdrawals are still fairly similar in interests to the high achievers in curriculum 2, so that some basis exists for believing that this group also might adapt well within curriculum 2.

The curriculum 2 high achievers exhibit interest patterns more similar to groups in curriculum 1 than to groups in curriculum 2. It might be expected then, that they might adapt well within curriculum 1.

On the basis of the assumption that interest patterns and success in a curriculum tend to go together, one might approve curriculum 1 low achievers transferring to curriculum 2 and with somewhat less confidence approve curriculum 1 withdrawals transferring to curriculum 2. Curriculum 2 high achievers might be approved to transfer to curriculum 1, although this too would be a somewhat tenuous approval.

Considering both abilities and interests, it seems probable that curriculum 1 low achievers would perform well and adjust adequately in curriculum 2. Curriculum 1 withdrawals would also be capable of performing well in curriculum 2, but the degree to which they might adjust adequately to the program is less certain. Finally, curriculum 2 high achievers would probably not be able to perform well in curriculum 1 because of lack of ability, although their interest patterns seem compatible with the program. It would seem desirable in these latter cases, for the purposes of effective selection, to consider transfers on an individual basis rather than on a group basis. The possibility of any success within curriculum 1 for curriculum 2 low achievers and withdrawals seems very remote.

Summary and Conclusions

Students enrolling in 1956 in the four-year engineering curriculum (called curriculum 1) at the Penn State main campus, were compared in academic achievement, aptitude, and interests to students enrolling at that same time in the two-year technical program in engineering at the university centers (called curriculum 2). It was found that the percentage of students within each curriculum exhibiting clear-cut academic success in their first semester was about the same. There was a significantly greater percentage of students withdrawing from curriculum 1, however. When members of each curriculum were categorized according to grade point average of two or above (high achievers), grade point average below two (low achievers), and withdrawal from the university, it was found that the lowest mean scores in ability tests obtained by any curriculum 1 subgroup exceeded the highest mean score obtained by any curriculum 2 subgroup. Furthermore, degree of success within a curriculum and ability scores seemed to be directly related.

When interests were considered according to the same categorizations, it was found that interests and degree of success seemed to be more related than interests and curriculum. The high achievers in both curriculums generally exhibited stronger interests in engineering and chemistry, than did either the low achievers or the withdrawals. Only five occupational areas were found, out of a total of 24 measured, in which any subgroup exhibited similar interests to people engaged in an occupation. These were engineering, chemistry, farmer, mathematics-science teacher and a composite area, group II.

It is felt that many of the high achievers in curriculum 2 probably would not fare so well academically within curriculum 1 because of lesser abilities, although their interests are like high achievers in curriculum 1. On the other hand, many of the curriculum 1 low achievers and possibly the cirriculum 1 withdrawals might do quite well within curriculum 2. The interests of the curriculum 1 low achievers seem more clearly pertinent to pursuit of an engineering career than do the interests of the curriculum 1 withdrawals, however. Accordingly it would seem wise to consider the merits of the individual case before recommending that a student withdrawing from the four-year curriculum pursue a two-year program.

FOOTNOTES

1. Grade Point Averages were computed by multiplying the number of credits for the course by the numerical value for the letter grade, where A = 4, B = 3, C = 2, D = 1, F = 0, and dividing by the total number of credits.

2. The 24 occupational categories and groups of categories were: medical, architecture, artist, group II (mathematics, engineer, chemist), farmer, math-science teacher, forest service, group V (YMCA physical director, personnel manager, social science teacher, minister), and group VIII (accountant, office man), group IX (life insurance), group X (advertiser, lawyer), and musician.

REFERENCES

1. Guthrie, L. T., Prediction of Academic Success for Freshmen at the Pennsylvania State College. Unpublished master's thesis, Pennsylvania State University, 1952.
2. Moore, B. F., and Castore, G. F. The Pennsylvania State University Academic Aptitude Examination. University Park, Pennsylvania: The Pennsylvania State University, 1947.
3. Strong, E. K., Manual for Vocational Interest Blank for Men. Stanford: Stanford University Press, 1951.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 1, September 1960)

AN OBJECTIVE AND EMPIRICAL STUDY OF THE EFFECTS OF LABORATORY EXPERIENCE IN A PROFESSIONAL EDUCATION COURSE PRIOR TO STUDENT TEACHING

DAN COX
Southern Illinois University, Carbondale, Illinois

"FIRST OF ALL the teacher's attitude will limit in a very real way her progress in learning to teach. It will determine the kinds of modifications that she is willing to attempt in her teaching, the energy with which the charges are pursued, and the learning that takes place (1)."

The attitude of the prospective teacher would seem to loom as a vital determinant in predicting the success of a teacher as attested by the foregoing statement.

Considerable research has been done to measure the effect of student teaching upon the attitude of teaching candidates during student teaching. These reports indicate that student teaching is probably in many respects the most important single course in the professional preparation of teachers. Since student teaching is largely a laboratory-type course and so significant in the education of teachers, it appears then that pre-student teaching courses with laboratory experiences would also improve a teacher's preparation. Therefore, the purpose of this research is to determine attitudinal changes of teaching candidates that resulted from laboratory experiences when these experiences are a feature of a pre-student teaching course, Human Growth and Development.

The Problem

Do the attitudes of prospective teachers change as a result of professional laboratory experience prior to student teaching? The aim is twofold: 1) to determine the direction and extent of changes of attitude toward children during the period of experimentation as measured by the Minnesota Teacher Attitude Inventory, and 2) to ascertain the relationship between these objectively measured attitudes and empirically measured appraisals as determined by the prospective teachers.

Sources of Data

The sample for this study was drawn from juniors and seniors at a southwestern state teachers college who were enrolled in a professional education course, Human Growth and Development. The majority of the students were between 20 and 22 years of age at the time of the study. One hundred and twenty-two cases were included in this study. The attitudes of the prospective teachers toward school work and toward pupils were measured objectively by the use of the Minnesota Teachers Attitude Inventory (MTAI) and the empirical device of personal appraisal was used.

Procedure

Objective Measure. The MTAI was administered to all students enrolled in the course, Human Growth and Development, at the beginning and again at the end of the course. The students were told that the MTAI scores would be kept in confidence and would not be used in grading. The students were randomly divided into control and experimental groups. The members of the experimental groups were assigned to children and youth organizations of the community. Each student worked with his group for a period of nine consecutive weeks. Organizations to which members of the experimental group were assigned were: Boy and Girl Scouts, Cub and Brownie Scouts, De Molay, Rainbow Girls, Saturday Morning Study Hour and Saturday Morning Playground. The members of the experimental group had course instruction plus the laboratory feature. The members of the control group received only the course of instruction. All sections of the class contained members from both control and experimental groups.*

Table I is a summary of significance of the

---

*Footnotes will be found at the end of the article.

TABLE I

A SUMMARY OF SIGNIFICANCE OF THE DIFFERENCES BETWEEN THE EXPERIMENTAL AND CONTROL GROUPS FROM PRE-TEST TO POST-TEST FOR FALL AND SPRING SEMESTERS ON TEST OF MINNESOTA TEACHER ATTITUDE INVENTORY

Group	Mean	S.D.	S.E. of Mean	S.E. Dif. of Means	t	t.01	t.05	Null Hyp.
Fall Exp.								
Pre-Test	29.18	28.95	4.64	3.37	6.98	2.71	2.02	Rej. at .01
Post-Test	52.70	31.61	5.07					
Fall Control								
Pre-Test	21.85	34.70	5.56	4.15	5.18	2.71	2.01	Rej. at .01 Rej. at .05
Post-Test	43.35	35.70	5.72					
Spring Exp.								
Pre-Test	34.71	29.90	6.69	9.14	5.73	2.84	2.09	Rej. at .01
Post-Test	55.05	27.85	6.23					
Spring Control								
Pre-Test	28.00	33.60	5.52	9.08	8.99	2.84	2.09	Rej. at .01
Post-Test	50.71	32.25	7.21					

TABLE II

TEST OF SIGNIFICANCE BETWEEN GAINS OF EXPERIMENTAL AND CONTROL GROUPS OF FALL SEMESTER ON MINNESOTA TEACHER ATTITUDE INVENTORY

Student No.	Experimental Gain ge	Control Gain ge	Student No.	Experimental Gain ge	Control Gain ge
1	5	2	21	21	36
2	22	17	22	14	-1
3	9	23	23	77	21
4	21	6	24	19	47
5	-17	24	25	18	20
6	29	12	26	24	23
7	57	68	27	28	25
8	12	45	28	14	38
9	-2	56	29	-12	35
10	-12	-15	30	27	36
11	36	18	31	0	13
12	31	52	32	45	-9
13	36	61	33	41	-7
14	27	53	34	73	44
15	33	1	35	17	7
16	16	-7	36	44	16
17	24	25	37	34	-10
18	14	-22	38	28	47
19	24	8	39	-21	12
20	51	49	40	36	57

Mge = 23.58
Mgc = 23.15

σge = 21.00
σgc = 22.93

σMge = 3.37
σMgc = 3.71

t = .086

TABLE III

TEST OF SIGNIFICANCE BETWEEN GAINS OF EXPERIMENTAL AND CONTROL GROUPS OF SPRING SEMESTER ON MINNESOTA TEACHER ATTITUDE INVENTORY

Student No.	Experimental Gain ge	Control Gain ge
1	32	0
2	-10	8
3	36	42
4	13	-11
5	35	-2
6	23	6
7	7	37
8	39	13
9	29	9
10	9	25
11	42	50
12	-7	11
13	13	35
14	41	30
15	28	55
16	44	29
17	33	34
18	10	33
19	-21	39
20	-17	28
21	48	6

Experimental gain = 427

Control gain = 477

Mge = 20.33
Mgc = 22.71
t = .39

$\sigma ge = 20.5$
$\sigma gc = 17.72$

$\sigma Mge = 4.59$
$\sigma Mgc = 3.96$

differences for the experimental and control groups from pre-test to post-test for the fall and spring semesters.

Table I indicates that both experimental and both control groups made statistically significant gains during the experimental period. The null hypothesis was rejected by all four groups at the .01 level of confidence. Therefore, it is assumed that this change in attitude can be attributed to the instruction.

Attitudinal Changes of Experimental Groups as Compared with Corresponding Control Groups

Objective Measure. Fall Semester. These data are treated in terms of mean gains. Table II compares the fall semester experimental group with the corresponding control group. The experimental group achieved a mean gain of 23.58 points on the MTAI from pre-test to post-test. The corresponding control group made a mean gain of 23.15. The experimental gain S. D. was 21. The control gain S.D. was 22.93. The S.E. of the mean gain of the control group was 3.71. The "t" test resulted in ratio of .086. Therefore, the significance of the experimental group over the control group was not significant at the .01 level of confidence.

Spring Semester. Table III reveals the results of the test of significance between the control and experimental groups for the spring semester. The results are based upon the gain scores made by the groups as a whole and are the results of the pre-test and post-test of the MTAI One may observe that the experimental group--the group whose members were assigned to professional laboratory experiences--earned a mean gain of 22.71. In an attempt to determine the significance of the differences between these two groups, it was found to be .39. This difference is not significant at either the .01 or the .05 level of confidence.

Empirical Measure. The students of the experimental groups completed a form with the following directions: "Describe your reactions to your experiences which you have had this semester in your work with your community activity assignment. Try to relate your remarks to your knowledge and understanding of boys and girls. It is not necessary that you sign your name." The participants were told that the written accounts they were about to prepare would not be read for a number of weeks and that the accounts were an attempt to evaluate the program.

The major purpose for asking for this type of information was so that the conclusions might be compared with the results of the objective measure mentioned above.

These subjective accounts were completed and gathered. Later they were read and in respect to subjective evaluation, it should be remembered that it is the total effect of the answers which are important rather than the specific aspects considered singly.

Sims (2) method was used for classifying the subjective accounts.

The participants of the study indicated that there was considerable personal value in their laboratory experiences. The extent of the value was varied. All the subjects did not find the greatest value in the same area. Several pointed out that the laboratory activity gave meaning to the theory discussed in class. Others seemed to feel that the actual experience of working with youngsters helped them to establish their present ability as a guide of children. Others pointed out that the experiences helped them to understand children and mentioned their enthusiasm for more work with children and youth.

Conclusions

The following conclusions were drawn from the data which were presented in this study:

Objective Instrument

1. The gain from pre-test to post-test was significant at the .01 percent level of confidence for all groups.
2. The experimental groups did not show a statistically significant gain over the corresponding control groups.

Empirical Instrument

1. The empirical measure strongly indicated the value of professional experience prior to student teaching in the preparation of teachers when the experiences are a requirement for a course related to human growth and development.

Empirical evidence strongly supports the thesis that the attitudes of prospective teachers toward teaching can be improved as a result of a professional education course, Human Growth and Development. The objective evidence did not support empirical evidence.

REFERENCES

1. Barr, A. S., and Rippen, N. D., "The Attitude of Teachers Toward Supervision," Journal of Experimental Education, June, 1935, p. 237.
2. Sims, V. M., "The Objectivity, Reliability, and Validity of an Essay Examination Graded by Rating," Journal of Educational Research, Vol. 24, October, 1931, pp. 216-223.

TABLE IV

SUMMARY OF THE DISTRIBUTION OF PARTICIPANTS' SUBJECTIVE ACCOUNTS AS TO THE VALUE OF THEIR LABORATORY EXPERIENCES

Group	Classification	Number	Percent
I	Superior Report	35	57.37
II	Average Report	24	39.34
III	Inferior Report	2	3.28
Total		61	99.99

FOOTNOTE

The author has chosen to use a cluster sample rather than a random sampling technique because of class size and time. It is believed that generalizations from this group's results to succeeding classes at this college can be made from the results of this experiment, hence inferential statistics are used in the treatment of the data here presented.

The students who participated in the experiment were not matched into pairs; therefore, the following formula was used to determine the effect of professional laboratory experiences upon their attitude toward teaching. The formula,

$$t = \frac{Mx - My}{\sqrt{\sigma_2 Mx + \sigma_2 My - 2r\, xy^{\sigma} Mx^{\sigma} My}}$$

was applied to the results of the MTAI which was used as the objective measure in this study. The above formula established the significance of the difference of means for pre-test to post-test for each group. It was then necessary to compare each experimental group's performance with the corresponding control group's performance for the data. For these data, the following formula was used:

$$t = \frac{Mge - Mgc}{\sqrt{\sigma_2 Mge + \sigma_2 Mgc}} .$$

The "t" obtained was interpreted in terms of the desired level of confidence (.01) through reference to Fisher's special table of "t".

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 1, September 1960)

A COMPARISON OF THREE WAYS OF PHRASING LIKERT-TYPE ATTITUDE ITEMS*

WILLARD E. NORTH
JOHN SCHMID**
University of Arkansas

Introduction

A NUMBER of years ago, Rundquist and Sletto (2) found that the phrasing of questions in attitude and personality scales affected the nature of responses. Their results suggested that reconsideration of the phraseology of items used in the Air Force for assessing job satisfaction was in order.

Examination of the attitude items revealed several inconsistencies in item phraseology. Of those variations detected, two types predominated. It was noted that some items were stated with a personal orientation for the respondee, while others were stated with an impersonal orientation. The second inconsistency in item construction was that some items contained a qualifying phrase while others did not.

As examples of how an attitude item might be stated with either a personal or impersonal orientation, consider the following two items: 1) "My supervisor is excellent;" and 2) "Air Force supervisors are excellent." In responding to the first item, the subject is required to consider only his own supervisor, while the second item requires the subject to consider the Air Force supervisors in general.

Examples of the second type of variation, i.e., items with and without qualifying phrases are given by the following two items: "My supervisor is excellent;" and "My supervisor is better than any other supervisor I might have had." The first item requires the respondee to agree or disagree with the statement about the categorical excellence of his supervisor. The second item permits the respondee to have a qualification built into his response. He must agree or disagree with the statement on the relative excellence of his supervisor when compared with other supervisors, but the subject need not give his supervisor his unqualified endorsement. Thus, it may be seen that a proposition intended to serve as an attitude item might be phrased in at least four ways as shown in the following example:

	Unqualified	Qualified
Personalized	My supervisor is excellent	My supervisor is better than other supervisors I might have had.
Impersonalized	Air Force supervisors are excellent.	Air Force supervisors are better than those found in civilian life.

Question was raised about the effect of the type of phrasing upon statistical characteristics, such as internal consistency, test-retest reliability, and independence of attitude areas. Because of the youthfulness of the airmen subjects, the Impersonalized-Qualified type of items were dropped from the study. The majority of subjects would have had very little opportunity to compare the Air Force job environment with that found in civilian life.

Hypotheses

Three hypotheses were developed, one for each of the three statistical characteristics to be examined in this study.

Hypothesis I: Personalized-Qualified items will produce the most internally consistent scales; Personalized-Unqualified items will produce less internally consistent scales; and Impersonalized-Unqualified items will yield the least internally consistent scales.

---

* This investigation was carried out under the Air Force Personnel and Training Research Center in support of Project No. 7950. Permission is granted for reproduction, translation, publication, and use or disposal in whole or in part by or for the United States Government.

**Both authors were with the Personnel Laboratory, WADC, when this study was conducted.

Hypothesis II: Personalized-Qualified items will produce scales with the highest test-retest reliability; Personalized-Unqualified items will produce scales that are less reliable; and Impersonalized-Unqualified items will yield scales with the lest test-retest reliability.

Hypothesis III: Impersonalized-Unqualified type items will demonstrate the highest inter-area correlations when measures on several attitude areas are obtained. Personalized-Unqualified type items will produce the second highest inter-area correlations. Personalized-Qualified type items will yield the lowest inter-area correlations. Low inter-area correlations are, of course, deemed important for diagnosticity.

Procedure

For this study five areas of job satisfaction or attitude derived from a previous factor analysis study were used (3). These were designated as: Supervisors, Air Force Bureaucracy, Co-workers, Living Conditions, and Job. For each of these five areas of job satisfaction, five-item scales representing the three types of item structure were developed. The same basic five items were used for each of the three scales, but were modified so that one scale contained the items written in a personalized-qualified fashion; a second scale contained the same items written in a personalized-unqualified fashion; and the third scale contained these items written in an impersonalized-unqualified fashion.

The above procedure resulted in a total of 75 items. In order to control for possible order effect, the items were partially randomized in such a fashion that each 15-item segment contained an item representing each of the three-item types from each of the five areas; i. e., each of the five areas of job satisfaction was measured by five items written with the personalized-qualified phraseology, five items written with the personalized-unqualified phraseology, and five items written with the impersonalized-unqualified phraseology. Items were assigned to each segment of the survey using a table of random numbers. A table of random numbers was also used to order the fifteen items within each segment. Making use of this partial randomization procedure, two forms were prepared. The only difference in the two forms was the order of item presentation. One form was labeled A; the other form was labeled B.

A total of 199 subjects was obtained. The subjects were enlisted personnel from one Strategic Air Command Air Force Base who were attending a two-week indoctrination school. Form A of the survey was administered on the first day of the school and Form B administered on the last day. A standard five response IBM answer sheet was used. The subjects were instructed to mark A if they strongly agreed with the statement, B if they agreed, C if they were uncertain, D if they disagreed, and E if they strongly disagreed.

It was decided to score item responses dichotomously rather than use the five-point scale. The strongly-agree and agree responses were combined and the disagree and strongly-disagree responses were also combined. The neutral or uncertain response was assigned to the unfavorable-attitude side of the dichotomy. The total score for each of the five scales was the number of favorable responses. Each scale, therefore, had a possible range of zero through five.

Results

Scale Means and Standard Deviations— The scale means are presented in Table I. Although it is not necessary to examine the means for evaluation of the three hypotheses, a rather interesting result does occur. For five areas, the subjects responded more favorably when using the Personalized-Unqualified item type than when using the other two. When scales, using the Personalized-Qualified and Impersonalized-Unqualified item types are compared, no consistent results are observed.

The scale standard deviations are presented in Table II. The standard deviations tend to be somewhat larger for the Personalized item types than for the Impersonalized item types. Between the two Personalized item types the differences are quite small and a number of inversions may be noted. Thus it would appear that any differences found in reliabilities for the different item type scales are probably attributable to factors other than differences in standard deviations.

Internal Consistency—Hypothesis I states that the Personalized-Qualified type of item would have the highest internal consistency and Personalized-Unqualified and Impersonalized-Unqualified item types would be decreasingly internally consistent in that order. To determine the consistency of each scale, the Kuder-Richardson Formula 20 was used. These results are presented in Table III.

It is seen that the Impersonalized-Unqualified item type scales have lowest internal consistency with the exception of Form A of the Job area. In the other four areas there appears to be no differences between the Personalized-Qualified and Personalized-Unqualified items. It appears that the data only partially supports Hypothesis I.

Test-Retest Reliabilities—Hypothesis II states that the Personalized-Qualified item type would demonstrate the highest test-retest reliability and that the Personalized-Unqualified and the Impersonalized-Unqualified item types would show decreasing reliabilities in that order. The Person-

TABLE I

JOB SATISFACTION SCALE MEANS BY AREA AND ITEM TYPE

	Supervisor		Air Force Bureaucracy		Co-workers		Living Conditions		Job	
Item Type	A*	B**	A	B	A	B	A	B	A	B
Personalized-Qualified	2.46	2.40	2.15	2.16	3.40	3.41	1.09	.91	2.72	2.83
Personalized-Unqualified	3.29	3.26	2.70	2.60	3.77	3.80	1.80	1.50	3.60	3.62
Impersonalized-Unqualified	2.70	2.89	2.11	2.31	2.72	3.03	1.29	1.00	3.30	3.53

* Initial Test
**Retest

TABLE II

JOB SATISFACTION SCALE STANDARD DEVIATIONS BY AREA AND ITEM TYPE

	Supervisor		Air Force Bureaucracy		Co-workers		Living Conditions		Job	
Item Type	A*	B**	A	B	A	B	A	B	A	B
Personalized-Qualified	1.78	1.87	1.83	2.03	1.68	1.69	1.38	1.40	1.54	1.54
Personalized-Unqualified	1.75	1.66	1.86	2.01	1.61	1.84	1.58	1.64	1.57	1.53
Impersonalized-Unqualified	1.17	1.18	1.73	1.91	1.63	1.75	1.32	1.28	1.58	1.44

* Initial Test
**Retest

TABLE III

INTERNAL CONSISTENCY OF JOB SATISFACTION SCALES BY AREA AND ITEM TYPE (KR 20)

Item Type	Supervisor		Air Force Bureaucracy		Co-workers		Living Conditions		Job	
	A*	B**	A	B	A	B	A	B	A	B
Personalized-Qualified	.781	.831	.807	.881	.796	.853	.697	.775	.642	.651
Personalized-Unqualified	.814	.784	.811	.873	.808	.861	.708	.775	.756	.743
Impersonalized-Unqualified	.509	.510	.777	.840	.714	.807	.594	.674	.711	.684

* Initial Test
**Retest

TABLE IV

TEST-RETEST RELIABILITIES OF JOB SATISFACTION SCALES BY AREA AND ITEM TYPE

Item Type	Supervisor	Air Force Bureaucracy	Co-workers	Living Conditions	Job
Personalized-Qualified	.781	.726	.773	.735	.711
Personalized-Unqualified	.787	.665	.778	.661	.684
Impersonalized-Unqualified	.577	.657	.668	.630	.648

TABLE V

INTERCORRELATIONS FOR THE FIVE JOB SATISFACTION AREAS

		Initial Test					Retest				
Area	Item Type	Super-visor	Air Force Bureauc-racy	Co-workers	Living Con-ditions	Job	Super-visor	Air Force Bureauc-racy	Co-workers	Living Con-ditions	Job
Supervisor	Pers. -Qual.		154	230	229	155	781	162	249	089	163
	Pers. -Unqual.		276	301	155	242	787	211	305	159	194
	Impers. -Unqual.		438	351	239	410	577	358	257	201	361
Air Force Bureaucracy	Pers. -Qual.			-008	122	146	173	726	047	170	147
	Pers. -Unqual.			144	187	274	218	665	064	253	221
	Impers. -Unqual.			193	147	359	322	657	080	220	341
Co-workers	Pers. -Qual.				318	287	225	-001	773	156	197
	Pers. -Unqual.				138	334	311	104	778	122	252
	Impers. -Unqual.				263	420	337	212	668	241	450
Living Conditions	Pers. -Qual.					126	224	107	287	735	090
	Pers. -Unqual.					086	115	245	137	661	052
	Impers. -Unqual.					177	181	121	172	630	154
Job	Pers. -Qual.						195	122	247	109	711
	Pers. -Unqual.						219	185	280	119	684
	Impers. -Unqual.						419	323	411	156	648
Supervisor	Pers. -Qual.							251	286	098	170
	Pers. -Unqual.							237	311	140	189
	Impers. -Unqual.							367	292	185	457
Air Force Bureaucracy	Pers. -Qual.								009	152	169
	Pers. -Unqual.								080	197	251
	Impers. -Unqual.								144	195	361
Co-workers	Pers. -Qual.									206	144
	Pers. -Unqual.									160	250
	Impers. -Unqual.									210	448
Living Conditions	Pers. -Qual.										118
	Pers. -Unqual.										106
	Impers. -Unqual.										220
Job	Pers. -Qual.										
	Pers. -Unqual.										
	Impers. -Unqual.										

TABLE VI

MEDIAN INTER-AREA CORRELATIONS FOR EACH ITEM TYPE

	Initial Test	Retest
Personalized-Qualified	.155	.161
Personalized-Unqualified	.259	.193
Impersonalized-Unqualified	.307	.256

alized-Qualified and the Personalized-Unqualified item types show approximately equal test-retest reliabilities for two attitude areas (see Table IV). For the other three areas, the Personalized-Unqualified item scales are slightly less reliable. The Impersonalized-Unqualified item type consistently shows the lowest test-retest reliabilities. Again the hypothesis seems to be partially supported.

Area Intercorrelations—Hypothesis III states that the five areas of job satisfaction would show the greatest independency for the Personalized-Qualified types, with the Personalized-Unqualified and Impersonalized types showing greater dependency in that order. Table V shows all the intercorrelations of the three item types for the five job satisfaction areas.

To examine Hypothesis III, median correlations of the five attitude areas for each type separately were computed. This was done for the initial testing and retesting separately (see Table VI). The median inter-area correlations suggests that Hypothesis III is substantiated.

Conclusions

The Personalized item type appears to be superior to the Impersonalized item type in terms of the statistical characteristics examined in this study. Regarding the two forms of the Personalized items, differences are not too conclusive for the Qualified and the Unqualified phrasing on internal consistency. However, on test-retest reliability and independency of attitude areas, the Qualified form seems superior. For purposes of diagnosticity, this is important (1). In general, the evidence suggests that if the attitude object lends itself to all three types of phraseology, it would seem that the Personalized-Qualified type of item phrasing is the best.

REFERENCES

1. Bennett, G. K. and Doppelt, J. E. "The Evaluation of Pairs of Tests for Guidance Purposes," Educational and Psychological Measurement, VIII (1948), pp. 319-25.
2. Rundquist, Edward A. and Sletto, Raymond F. Personality in the Depression: A Study in the Measurement of Attitudes, Minneapolis: University of Minnesota, 1936.
3. Schmid, John, and others, Analysis of Job Satisfaction, AFPTRC Research Report, TN-57-30, March 1957.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 1, September 1960)

MOTIVES IN SELECTING ELEMENTARY AND SECONDARY SCHOOL TEACHING*

GERHARD LANG
Fairleigh Dickinson University, Teaneck, New Jersey

SEVERAL STUDIES have sought to determine motives for teaching. Recently, Lloyd-Jones (14), Chaltas (3), and Lang (11) have surveyed the relevant previous research. Interviews (8, 16), questionnaires (1, 5, 8, 17, 19, 22, 31), and essays on the theme "Why I want to teach" (7, 32) were the devices most frequently employed. Generally, in view of the critical teacher shortage, the direction of these studies was focused on finding answers to immediate problems, rather than on testing hypotheses derived from theory.

Theories of vocational choice, the concept of the multi-dimensionality of the teacher role, and studies of teacher personality provide a theoretical framework for research in this area.

Personality is considered to be an important determinant of vocational choice (21, 24), which is, to a large extent, an expression of basic psychological needs (6). Studies have shown that there are personality characteristics which differentiate between (20) as well as within (10, 15) professional groups.

It is generally recognized that the modern teacher plays many roles inside the classroom (18, 23). At the elementary and at the secondary level, some roles more than others tends to be emphasized. For instance, the elementary teacher confronts her pupils predominantly as a mother-substitute, whereas the secondary teacher, in her role of subject-matter specialist, functions primarily as director of learning. It seems reasonable to assume that one of the important reasons why women select one rather than the other level of teaching is the fact that they like the activities associated with the teaching role appropriate to that level.

The findings of studies of teachers' personalities (15, 25-27) and teachers' classroom behavior (28, 29) suggest that teaching serves as a distinct outlet for certain psychological needs and that these may differ for individuals who elect to teach at the elementary or secondary level.

The purpose of this study was to determine to what extent these differences, if any exist, are reflected in the motives for choosing the teaching profession. Motives for teaching were operationally defined as : (a) reasons comprising the Lang Scale of Motives for Teaching and (b) psychological needs measured by the Edwards Personal Preference Schedule.

Two hypotheses were tested: (a) elementary teachers will emphasize motives relating to the mothering aspects of teaching, (b) secondary teachers will emphasize motives relating to the academic and intellectual facets of teaching.

Subjects, Instruments

The subjects of this study were 101 female elementary and 87 female secondary teachers, 21-29 years of age, and currently teaching in 16 New York City public schools. Their experience ranged from one month to 10 years.

The Lang Scale of Motives for Teaching (12) and the Edwards Personal Preference Schedule (4) were the data-gathering devices of this study.

The Lang Scale measures the degree to which (i. e. "least", "slightly", "half/half", "much", "most") each of 25 reasons has influenced a subject to choose teaching. Checklists used in previous investigations listed reasons such as "I like working with children" which may be considered to be generally admired and socially desirable. In constructing the Lang scale, an effort was made to include reasons, such as "I like to be in a position of authority", which evoke varying degrees of social approval. Judgments of 60 people, 18 educators and 42 non-educators were used in determining the social acceptability values of each of 76 potential reasons for teaching, in accordance with Thurstone's method of equal appearing intervals (30). The bases for selecting 25 reasons for inclusion in the scale were (a) the assumed appropriateness of the reason for either the elementary or secondary level of teaching and (b) the usual criteria governing the development of a Thurstone-type scale (30).

---

* Based, in part, on a doctoral dissertation (11) submitted to Columbia University in partial fulfillment of the requirements for the Degree of Doctor of Philosophy.

The Edwards Personal Preference Schedule, consisting of 210 pairs of statements, is designed to measure the relative importance of 15 psychological needs based on the subjects' self-descriptions. Each pair of statements is matched approximately for mean social desirability value, as previously judged by 86 men and 66 women, in order to minimize the effect of social desirability on item choice.

Procedure

The subjects were asked to complete a brief background questionnaire and then to consider the degree to which each of 25 reasons has influenced their choice of teaching as their profession. They were asked to check three reasons which were most influential in their choice, three reasons which were least influential, five reasons which were slightly, and five reasons which were much influential. The nine remaining reasons were those considered to be about half/half influential in their decision to teach. The EPPS was administered with the standard instructions printed on the test booklets. The data of subjects who incorrectly responded to the Scale of Motives and/or received a consistency score of less than 10 on the EPPS were not used. The two instruments were responded to anonymously.

Results

1. Both elementary and secondary school teachers rated as most influential reasons which are socially acceptable; they rated as least influential reasons which are socially unacceptable. All rank difference correlations obtained between social acceptability values of the reasons and frequency of checkings by subjects in the two response categories were significant at better than the .001 level of significance.
2. Elementary teachers emphasized the reasons "I like working with children" ($X^2 = 14.86$, $p < .001$) and "I like to give to and to receive love from children" ($X^2 = 12.97$, $p < .005$).
3. Secondary teachers emphasized the reasons "I like the continuous opportunity to learn" ($X^2 = 6.77$, $p < .05$) and "I enjoy the intellectual fellowship of other teachers" ($X^2 = 20.11$, $p < .001$).
4. Elementary teachers manifested a greater n·Nurturance ($r = .23$, $p < .001$) and lesser n·Achievement ($r = .12$, $p < .05$) than did secondary teachers.
5. None of the other motives differentiated significantly between the two groups of teachers.

Discussion

Findings of past research implied that, generally, teachers state reasons for teaching which might be considered socially acceptable. However, the validity of this deduction has not been shown. Previous investigators were not concerned with social acceptability per se as a factor to be studied. Generally, their subjects had not been given the opportunity to respond anonymously, a fact which may have been instrumental in their volunteering primarily socially desirable reasons. Furthermore, in most studies, teachers were confronted with a limited array of reasons, most of which were of a socially acceptable nature. In this study, the subjects responded anonymously to a scale comprising a wide variety of reasons which were selected on the basis of psychological theory. Social acceptability values were accurately determined and not arbitrarily inferred.

The results show that teachers, even when promised anonymity, will emphasize socially acceptable reasons for selecting teaching. In this respect, elementary and secondary teachers are alike. Apparently, teachers are either reluctant and/or unable to admit the powerful sway of motives which are socially less desirable. In private communications, teachers will agree that teaching is a convenient job for a woman. Yet only 19 teachers (10%) regarded this reason as most influential in their choice of teaching despite the fact that this motive ranked in the middle of the social acceptability continuum. On the other hand, the reason "I like working with children", which ranked highest on the continuum was checked by 116 teachers (62%) as being most influential.

The avoidance of less desirable reasons is probably a reflection of the conforming behavior which our society imposes on teachers, who, in their role as surrogates of middle class morality, may be reluctant to admit to motives which are at variance with the established values of society (2, 13).

Preference for the socially desirable reasons may be also attributable to the fact that most of the subjects may have been unable to acknowledge the potent influence of some of the less desirable reasons in their choice of teaching. They may, for instance, not be able to recognize that a need to express hostility is, unconsciously, responsible for their vocational choice. And yet, Jersild (9), and Symonds (29) have noted that the expression of hostility in the classroom is a widespread phenomenon which serves as a defense against the teacher's own anxiety.

In a recent study, Jersild (9) has elicited a variety of additional intrapsychic problems experienced by teachers, such as loneliness, the search for life's meaning, sexual conflicts, and lack of the freedom to feel. The majority of teachers who responded to his questionnaire accepted the idea that a major goal in education should be to help young people to understand and accept themselves. Furthermore, the vast majority of teachers saw the necessity of their growing in self-understanding and welcomed the opportunity to discuss personal and emotional issues confronting them in workshops, special courses, and group therapy. The process of self-discovery on the part of teachers involves, among other things, a recognition of their motives for teaching and an understanding of their personality needs. Even though certain motives may appear to be somewhat far fetched or repugnant, they may nevertheless be strong factors in their choice of a teaching career.

The Scale of Motives for Teaching which has served as a research instrument may be useful as a tool in counseling pre-service as well as in-service teachers. In discussing their responses to this scale, teachers may be helped to "face themselves" by examining their motives for teaching and, in doing so, may better understand their personality needs.

The hypothesized differences in motives between female elementary and secondary teachers were supported. The findings suggest that possibly one important reason why female elementary teachers are attracted to teaching is that they like the activities associated with the teacher role of mother-substitute. They seem to like the companionship of children and to desire giving and receiving love from them. These teachers, apparently, like to treat others with kindness, like to be generous with others, and desire to have others confide in them about personal problems, characteristics which are associated with n· Nurturance.

With regard to female secondary teachers, the data suggest that they see in teaching an opportunity to direct the learning of others, to pursue their own quest for knowledge, and to enjoy the companionship of intellectually stimulating fellow teachers. Apparently, they like to be successful, desire to be regarded as an authority, attempt to accomplish something of great significance, and strive to do things better than others, all characteristics indicative of n· Achievement.

The apparent attraction of elementary teachers to the emotional and that of the secondary teachers to the intellectual facets of teaching does not imply that the latter do not want to establish warm personal ties with others. Teachers at both levels like to be with people and do not differ significantly in n· Affiliation. However, female elementary teachers apparently like to relate to people who are relatively immature, who are receptive to love, and who are willing to reciprocate. Female secondary teachers, on the other hand, seem to prefer people who are relatively mature and who, on a reciprocal basis, can serve as outlets for their intellectual needs.

The finding of statistically significant differences in motives between female elementary and secondary teachers should not conceal the fact that there is considerable variability within each teaching level. Furthermore, in view of the selective nature of the sample studied—New York teachers—inferences to the population of female American teachers may not be warranted. Comparable studies, drawing upon samples in different localities throughout the country, are indicated.

Summary

It has been suggested that teaching satisfies certain psychological needs and that these may differ for individuals who elect to teach at the elementary or secondary level. The present study sought to determine to what extent these differences, if any, are reflected in motives for choosing the teaching profession. The Lang scale of Motives for Teaching and the EPPS were administered to 101 female elementary and 87 female secondary teachers, 21-29 years of age, teaching in 16 New York City public schools.

Motives for teaching emphasized by elementary teachers related to the "mothering" role; those stressed by secondary teachers involved the "director of learning" role. Elementary teachers manifested a greater n· Nurturance and lesser n· Achievement than did secondary teachers. The two groups of teachers were similar with regard to the social acceptability of reasons perceived as influential in their choice of teaching.

The data support the view that female elementary and secondary teachers differ in some aspects of their personality. However, the differences found in the present study, though statistically significant, are not large enough to warrant a broad generalization.

BIBLIOGRAPHY

1. Best, J. W. "A Study of Certain Selected Factors Underlying the Choice of Teaching as a Profession," Journal of Experimental Education, 1948, 17, 201-259.
2. Brookover, W. B. A Sociology of Education, New York: American Book Co., 1955.
3. Chaltas, J. G. Factors Influencing the Choice Regarding a Teaching Career: a Review and Analysis of the Literature, unpublished Doctor's Dissertation, Columbia University, 1957.
4. Edwards, A. L. Manual of the Edwards Personal Preference Schedule, New York: Psychological Corp., 1954.
5. Fielstra, C. "An Analysis of Factors Influencing the Decision to Become a Teacher," Journal of Educational Research, 1955, 48, 659-667.

6. Forer, B. R. "Personality Factors in Occupational Choice," Educational and Psychological Measurement, 1953, 13, 361-366.
7. Hartford, E. F. "Why Two Hundred Chose Teaching," Phi Delta Kappan, 1948, 30, 126-127.
8. Hoyt, Mildred D. "Factors Which Influenced Pre-service Teachers' Choice of Elementary or Secondary Division," Dissertation Abstracts, 1955, 15, 1793 (Abstract).
9. Jersild, A. T. When Teachers Face Themselves, New York: Bureau of Publications, Teachers College, Columbia University, 1955.
10. Kelley, E. L. and Fiske, D. W. The Prediction of Performance in Clinical Psychology, Ann Arbor: University of Michigan Press, 1951.
11. Lang, G. Motives in Selecting Elementary and Secondary School Teaching, unpublished Doctor's Dissertation, Columbia University, 1958.
12. Lang, G. "An Experimental Scale to Measure Motives for Teaching," Journal of Educational Research, May, 1958, 687-694.
13. Lieberman, M. Education as a Profession, Englewood Cliffs, N. J.: Prentice-Hall, 1956.
14. Lloyd-Jones, Esther, and Holman, Mary V. "Why People Become Teachers," in L. J. Stiles (Ed.) The Teacher's Role in American Society, Yearbook of the John Dewey Society, New York: Harper, 1957.
15. Morey, E. A. "Vocational Interests and Personality Characteristics of Women Teachers," Australian Journal of Psychology, 1949, 1, 26-37.
16. Norton, J. L. "General Motives and Influences in Vocational Development," Journal of Genetic Psychology, 1953, 82, 263-278.
17. Ostlie, S. "Motivation for Entering Teaching," Journal of Teacher Education, 1956, 7, 80-81.
18. Redl, F., and Wattenburg, W. M. Mental Hygiene in Teaching, New York: Harcourt-Brace, 1951.
19. Ringness, T. A. "Relationship Between Certain Attitudes Toward Teaching and Teaching Success," Journal of Experimental Education, 1952, 21, 1-55.
20. Roe, Anne. "A Psychological Study of Eminent Psychologists and Anthropologists, and a Comparison with Biological and Physical Scientists," Psychological Monographs, 1953, 67, No. 2 (Whole No. 352).
21. Roe, Anne. The Psychology of Occupations, New York: Wiley, 1956.
22. Stewart, L. H. "Certain Factors Related to the Occupational Choices of a Group of Experienced Teachers," Peabody Journal of Education, 1956, 33, 235-239.
23. Stiles, L. J. (Ed.) The Teacher's Role in American Society, Yearbook of the John Dewey Society, New York: Harper, 1957.
24. Super, D. E. "A Theory of Vocational Development," American Psychologist, 1953, 8, 185-190.
25. Symonds, P. M. "Personality Adjustment of Women Teachers," American Journal of Orthopsychiatry, 1941, 11, 14-20.
26. Symonds, P. M. "Needs of Teachers as Shown in Autobiographies," Journal of Educational Research, 1943, 36, 662-677.
27. Symonds, P. M. "Needs of Teachers as Shown in Autobiographies," Journal of Educational Research, 1944, 37, 641-655.
28. Symonds, P. M. "Teaching as a Function of the Teacher's Personality," Journal of Teacher Education, 1954, 5, 79-83.
29. Symonds, P. M. "Characteristics of the Effective Teacher Based on Pupil Evaluation," Journal of Experimental Education, 1955, 23, 289-310.
30. Thurstone, L. L., and Chave, E. J. The Measurement of Attitude, Chicago: University of Chicago Press, 1929.
31. Wellbank, H. L. "Why Teach?" Journal of Educational Research, 1955, 48, 699-703.
32. Willcox, I., and Beigl, H. G. "Motivations in the Choice of Teaching," Journal of Teacher Education, 1953, 4, 106-109.

The Journal of Experimental Education

PUBLISHED QUARTERLY

Copyright 1960 by

DEMBAR PUBLICATIONS, INC.
Madison 3, Wisconsin

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 2, December 1960)

CHILDREN'S PERCEPTIONS OF THEIR TEACHERS' FEELINGS TOWARD THEM RELATED TO SELF-PERCEPTION, SCHOOL ACHIEVEMENT AND BEHAVIOR*

HELEN H. DAVIDSON and GERHARD LANG
City College, New York

Introduction

THE CHILD'S self-concept arises and develops in an interpersonal setting (30). Feelings about the self are established early in life and are modified by subsequent experiences. Among the significant people believed to affect the child's feelings about himself are first, his parents, and, later, his teachers. Ausubel (2) and Jourard and Remy (16) are among the few investigators who have reported results which support these theoretical contentions.

Rogers (24), Snygg and Combs (27), among others, assign the self-concept a central place in their personality theories and suggest that the individual's self-concept is a major factor influencing his behavior. Vigorous research in this area by Martire (17) and Steiner (28) has produced corroborative evidence for these views.

Only recently has the concept of the self been introduced into the school setting. Typical studies are those by Jersild (15), Reeder (23) and Stevens (29). Jersild demonstrated the value of the self-concept theory in making the educative process more valuable. Reeder, using grade school children and Stevens, working with college students, explored the relation between self-concept and school achievement. Both of these investigators found that positive feelings about the self are associated with good academic achievement.

A series of studies dealing with teacher-pupil relations have sought to determine a) how children see and feel about their teachers (11); b) how teachers see and feel about their pupils (5, 20); and c) how teachers think their pupils see themselves (22).

It has been widely recognized that teachers influence the personality development of their pupils (21). Perkins, for example, found that teachers who had completed several years of child study were able to promote healthier personality growth in children, defined in terms of congruency between the self and the ideal self. For this reason, many researchers, among them, Barr and Jones (3) and Symonds (31) are engaged in the study of personality development of the teacher herself.

Despite the abundance of research on these aspects of the school setting, an important dimension, not previously investigated, is how the child perceives his teacher's feelings toward him. In an investigation of this interaction, we not only may gain insight into the question of what qualities make for an effective teacher but also an understanding of how the child's perception of his teacher's feelings, irrespective of its accuracy, relates to his self concept, school achievement and classroom behavior.

It is the purpose of this investigation to determine what the relation is between children's perception of their teachers' feelings toward them and the variables: self-perception, academic achievement, and classroom behavior.

Specifically, three hypotheses were tested:

1. There exists a positive correlation between children's perception of their teachers' feelings toward them and children's perception of themselves. In behavioral terms it is predicted that the more favorable the child's perception of himself, the more positive will be his perception of teachers' feelings toward him.
2. There exists a positive relationship between favorable perception of teachers' feelings and good academic achievement.
3. There exists a positive relationship between favorable perception of teachers' feelings and desirable classroom behavior.

---

*This study was supported by a grant from the James McKeen Cattell Fund.

The Instrument

To test the hypotheses proposed, it was necessary to develop an instrument to measure self perception and the perception of the feelings of others. It was decided to use an adjective checking method, since it is direct and simple. Adjective check lists have been used to measure adjustment (18), self acceptance (4), empathy (9), character traits (13), and to distinguish the self perceptions of persons classified according to some social and psychological variables (26). In the main, these lists have been used with adults.

In developing the check list with children, words and phrases to be included were selected on the basis of the following three criteria:

1. The words should be those commonly used to describe how people feel toward and how people think of others, especially how teachers feel toward and think of children. An attempt was made to cover varied aspects of behavior and personality. For this purpose, lists already developed, like those of Allport (1), Gough (12), and Hartshorne and May (13), were scanned for appropriate words.
2. The words should be easy enough for children in approximately the 10-16 year age range to read and comprehend. The Thorndike-Lorge Frequency Count (33) was used to eliminate words which would be too difficult.
3. The list should contain about an equal number of words connoting positive and negative feelings.

From an initial pool of 200 trait names, 135 remained after the application of criteria 1 and 2. The next step was to determine the feeling tone of the 135 words. Each of the words was then rated by 35 teachers and 50 junior high-school pupils as favorable, unfavorable, or neutral. Only those words were retained which were judged by more than 80% of the teachers and 80% of the pupils as being favorable or unfavorable. The words judged neutral were eliminated.

Fifty words remained after the teachers and students judged them as favorable or unfavorable. The 35 words finally used are listed below along with the F or U rating received. Fifteen words were dropped either because of the level of difficulty or because of some duplication in meaning.

Fair	(F)	A hard worker	(F)
A nuisance	(U)	Bad	(U)
Afraid	(U)	A good sport	(F)
Cheerful	(F)	Considerate	(F)
A time waster	(U)	Not eager to study	(U)
Neat	(F)	Helpful	(F)
Not eager to learn	(U)	Careless	(U)
A leader	(F)	Sociable	(F)
Unhappy	(U)	Clever	(F)
Loving	(F)	Not alert	(U)
Outstanding	(F)	Smart	(F)
Loud	(U)	Silly	(U)
Generous	(F)	Kind	(F)
Nervous	(U)	Shy	(U)
Sensible	(F)	A sloppy worker	(U)
Polite	(F)	Dependable	(F)
Lazy	(U)	A day dreamer	(U)
Forgetful	(U)		

Administration and Scoring of the Check List—The children are instructed to decide how the teacher feels toward them with respect to each trait name, and then to rate it on a three-point rating scale: most of the time, half of the time, seldom or almost never. A favorable word is assigned a score of 3 when it is checked in the most of the time column; a score of 2 for half of the time, and 1 for seldom or almost never. For an unfavorable word the scoring is reversed.

The total score, the Index of Favorability, is obtained by adding the scores of all the words and dividing the total by the number of words checked. The higher the index, the more favorable is the child's perception of the teacher's feelings toward him. Theoretically, the index can range from 1.00 to 3.00.

Reliability and Validity—The Checklist of Trait Names was administered twice to four classes comprising 105 junior high-school children. The interval between the two administrations was from four to six weeks. A correlation of .85 was obtained (rank difference, $p < .001$).

The Checklist may be considered to have logical validity. However, it was desired to obtain a measure of empirical and concurrent validity. This was done by correlating the child's own perception of his teacher's appraisal of him with his classmates' perceptions of the teachers' feelings toward him. For this purpose, a modified version of the de Groat and Thompson Teacher Approval and Disapproval Scale (7) was administered along with Checklist to 93 children (3 classes). The de Groat and Thompson scale, as modified, consisted of 8 positive statements, such as, "Here is someone whom the teacher praises for trying hard," and 8 negative statements, such as, "Here is someone whom the teacher often points out as wasting too much time." For each statement, pupils were asked to name one to four of their classmates to whom these characteristics applied. They could also name themselves, if they so desired. Of the 93 children, 56 received 5 or more votes on one of the teacher approval and disapproval statements. For these 56 children, a teacher approval score was determined by subtracting the number of unfavorable statements on which five or more votes were received from the number of favorable statements on which five or more votes were received. A correlation of .51 was obtained (rank difference, $p<.001$) between the Index of Favorability and the teacher approval score.

The Checklist developed to assess children's perception of their teachers' feelings toward them appears to have satisfactory reliability and validity. Although the estimate of reliability and validity was based on a sample of junior high-school students, the list was considered appropriate also for the upper grades of the elementary school because of the way the words were chosen.

Experimental Design

Subjects—The subjects of this study were 89 boys and 114 girls, attending 4th, 5th, and 6th grades of a New York City public school. These children were distributed in 10 different classrooms. In terms of reading ability, the classes selected were in the upper half of their respective grade level. Originally, it was planned to test all 4th, 5th, and 6th grade children, but after preliminary experimentation, it was found that several words were too difficult for children of limited language ability. It was therefore decided to test children in those classes which were known to have the better readers.

The children represented a wide range in socio-economic status. It was possible to divide them into three distinct groups on the basis of their fathers' and mothers' occupation. The upper group, consisting of 63 children, came from families of professional people, white collar workers and business men; the middle social class group of 57 children had parents who were skilled workers, policemen and firemen; the low group contained 83 children of semi-skilled and unskilled workers and a number of unemployed.

Table I presents the background information for the 203 children involved in the study.

Procedure—The Checklist of Trait Names was administered twice to the children. At the first administration, the children were instructed to respond to the 35 adjectives comprising the list in terms of "My teacher thinks I am," and at the second testing, in terms of "I think I am." The first testing was done in the morning, the second in the afternoon. The "My teacher thinks I am" scale yields a measure of perceived teacher feelings, referred to henceforth as the Index of Favorability; the "I think I am" scale yields a measure of self-perception.

The teachers, nine women and one man, rated their pupils on academic achievement, on a four-point scale: Very Well, Adequately, Below Average, and Very Poorly. In the analysis of data, the last two categories were combined due to the paucity of cases in the category Very Poorly. At the same time, the teacher also rated each child on 10 behavioral or personality characteristics. A weight of +1 was assigned to each of the traits judged to be desirable. The four desirable traits were: eager, obedient, cooperative, assertive. A weight of -1 was given to the characteristics judged to be undesirable: disorderly, destructive, hostile, defiant, unfriendly, and troublesome. The sum of the weights yielded a behavior rating score ranging theoretically, from +4 (very desirable) to -6 (very undesirable). Subjects who received the 0 and minus behavioral ratings were combined into one group due to the small number of cases in these categories.

Results and Discussion

Hypothesis 1—There exists a positive correlation between children's perception of their teachers' feelings toward them and children's perception of themselves.

The two perceptual favorability indexes correlated .82 (product-moment, $p < .001$). The children who had a more favorable or a more adequate self-concept, that is, those who achieved a higher self-perception score also perceived their teachers' feelings toward them more favorably.

The finding of a significant correlation between the two kinds of perception lends support to the view that a child's assessment of himself is related to the assessment "significant people" make of him (30). In two previous research investigations, a close relationship was found between self-appraisal and children's perception of their parents' feelings toward them (2, 16). The present study for the first time has shown that a child's self-appraisal is significantly related to his perception of his teacher's feelings as well. Such a finding was anticipated in view of the fact that one role of the teacher, at least at the elementary level, is that of a "parent substitute." Several interesting questions may be raised: To what extent does a child's perception of his teacher's feelings resemble his perception of his mother's or father's feelings toward him? Does the child's perception of his present teacher differ from his perception of his previous teacher? Does favorability or perception decrease or increase with years in school?

Hypothesis 2—There exists a positive relationship between favorable perception of teachers' feelings and academic achievement. Table II presents the mean favorability scores and their standard deviations for the three levels of estimated achievement. The *F* ratio of 15.61 was significant at less than the .001 level. The three t tests were also significant at better than the .01 level.

Hypothesis 3—There exists a positive relationship between favorable perception of teachers' feelings and desirable classroom behavior. The findings pertaining to the relationship between children's perception and their classroom behavior are shown in Table III.

The overall *F* ratio of 7.38 was significant at less than the .001 level. The only significant t tests were those between the lowest category (0 and less) and all the other categories. In other words, the children who were rated as being disorderly,

TABLE I

DISTRIBUTION OF SUBJECTS IN THE TEN CLASSROOMS BY SEX AND SOCIAL CLASS STATUS

	Socio-Economic Status							
	Upper		Middle		Lower		Total	
Classroom	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
4-1	8	5	3	8	-	2	11	15
5-1	1	-	-	2	7	8	8	10
5-2	-	-	-	-	3	5	3	5
5-3	7	10	7	2	-	1	14	13
5-4	-	2	3	1	9	8	12	11
5-5	2	4	6	3	3	7	11	14
6-1	5	5	4	3	-	-	9	8
6-2	2	5	4	2	1	6	7	13
6-3	3	3	5	3	-	9	8	15
6-4	-	1	-	1	6	8	6	10
Total	28	35	32	25	29	54	89	114

TABLE II

INDEX OF FAVORABILITY AS RELATED TO THREE LEVELS OF ESTIMATED ACHIEVEMENT

	Achievement Category		
	Very Well	Adequately	Below Average
Mean Favorability Score	2. 68* (N=53)	2. 57 (N=111)	2. 40 (N=39)
S. D.	. 22	. 24	. 25

*The higher the score, the more favorable the child's perception of his teacher's feelings toward him.

TABLE III

INDEX OF FAVORABILITY AS RELATED TO FIVE LEVELS OF RATED BEHAVIOR

	Behavior Rating Category				
	Very Desirable		Desirable		Undesirable (0 and Minus Scores)
	(+4	+3)	(+2	+1)	
Mean Favorability Score	2. 62* (N=40)	2. 65 (N=54)	2. 58 (N=46)	2. 53 (N=23)	2. 39 (N=40)
S. D.	. 26	. 19	. 27	. 27	. 28

*The higher the score, the more favorable the child's perception of his teacher's feelings toward him.

defiant, unfriendly, or troublesome, perceived their teachers' feelings toward them as being less favorable than the children who were rated as being eager, cooperative, assertive and the like.

One of the axioms of educational psychology is the statement that a child learns only when he is motivated to learn. Furthermore, the basic incentives which a teacher can furnish are her acceptance of the child on the one hand, and approval on the other. The findings of the present study furnish supporting evidence. The teacher's feelings of acceptance and approval are communicated to the child and perceived by him as positive appraisals. It is likely that these appraisals encourage the child to seek further teacher approval by achieving well and behaving in a manner acceptable to his teacher. We may also begin this cycle with the child's behavior. The child who achieves well and behaves satisfactorily is bound to please his teacher. She, in turn, communicates positive feelings toward the child, thus reinforcing his desire to be a good pupil. Which of these variables serves as the primary determiner is a fact difficult to ascertain. It seems rather that they reinforce each other. The implication is clear. It is essential that teachers communicate positive feelings to their children and thus not only strengthen their positive self-appraisals but stimulate their growth, academically as well as interpersonally.

It should be emphasized that these findings do not imply causality but rather suggest that certain pupil characteristics, such as self-perception, perceived teacher feelings, achievement and behavior in school are interrelated.

In addition to the results relevant to the tested hypotheses, other findings will now be reported.

Sex Differences—Sex differences were observed with regard to the three variables studied: index of favorability, * achievement, and behavior in school. Girls perceived their teachers' feelings toward them more favorably than did the boys (girls mean = 2.60; boys mean = 2.52; $t = 2.41$, $p < .02$). The behavior ratings of the girls were more favorable than those of the boys ($x^2 = 10.72$, df = 4, $p < .05$); the girls were likewise rated more favorably in achievement, although this difference was not significant ($x^2 = 3.41$, df = 2, $.10 < p < .20$).

Past research has consistently shown that teachers report more problem behavior among boys (32). One explanation, though not widely accepted, is that boys are naturally more aggressive. Another view, more plausible, holds that our society encourages aggressive behavior in men (and men to be) and submissive behavior in women. Teachers, most of whom are women, especially in the primary grades, therefore regard boys' classroom behavior as disturbingly different from the norms of behavior appropriate to their own female sex. The temptation is great to reward children of one's own sex. Meyer and Thompson's study (19) is pertinent here. Teacher-pupil interaction of sixth-grade pupils were studied over a one-year period and analyzed in terms of "approval" and "disapproval" contacts. In addition, children were asked to nominate by the "Guess Who" technique which of their classmates receive their teacher's approval and disapproval. Both approaches yielded the same finding. Classroom observers, as well as the children themselves, noted that teachers expressed greater approval of girls and greater disapproval of boys. The findings of the present investigation, which ascertained directly children's perceptions of their teachers' feelings, are in accord with the results of prior research. The suggestion has been frequently made that more men should be urged to teach at the primary level. Findings such as those discussed above suggest the urgency to establish a sexual balance in the teaching staff at the primary grades. Not only is it desirable for boys to have a male model with whom to identify, but conditions may then be created which may assure greater teacher approval for boys and reduce teacher disapproval for behavior which is, to a large extent, culturally instigated.

Social Class Differences—Because of the distinct differences found in social class status in this group of children, it was decided to investigate the relation of social class to the index of favorability, achievement and behavior in school. All three variables are related to social class in the direction one would predict. These data are shown in Table IV.

It may be observed from Table IV that there is a decline in mean favorability index from the upper to the lower social class. Two of the three t tests were significant at better than the .01 level; t was not significant between the upper and middle social class groups. Children in the two advantaged social class groups perceive their teachers' feelings toward them more favorably than do the children in the lower class group.

Social class and achievement in school are significantly related ($x^2 = 18.38$, 4df, $p < .01$). The differences in the percentage of children in the several categories may be pointed out, especially the difference between the two extremes: in the upper social class 43% of the children were rated by their teachers as doing very well in school while only 15% were rated as doing below average work.

Social class and behavior in school as rated by

---

*The index used in this and subsequent analyses is based on the Check List score of the child's perception of his teacher's feelings toward him.

TABLE IV

SOCIAL CLASS STATUS RELATED TO FAVORABILITY INDEX, ACHIEVEMENT AND BEHAVIOR

	Upper Social Class (N=63)	Middle Social Class (N=57)	Lower Social Class (N=83)
Mean Favorability Index	2. 63	2. 60	2. 49
S. D.	. 26	. 22	. 26
Achievement Rating Category:			
Very Well (N=53)	43%*	34%	23%
Adequately (N=111)	31%	22%	47%
Below Average (N=39)	15%	36%	49%
Behavioral Rating Category:			
Very Desirable (N=94)	41%	29%	30%
Desirable (N=69)	23%	30%	46%
Undesirable (N=40)	20%	22%	58%

*These percentages are based on the N's of the Achievement and Behavior categories.

the teachers were not significantly related ($x^2 = 14.97$, 8df, $.05 < p < .10$). However, the distribution of children in the several categories reveal interesting differences. While the great majority of the children in the group were rated favorably by their teachers, there were 58% of the children in the lower class whose behavior was rated as undesirable while only 20% of the upper class children were so rated.

It has been suggested that teachers, as surrogates of middle class values, tend to give preferential treatment to the middle and upper social economic class pupils, and to withhold rewards from pupils who belong to the lower socio-economic class (6, 8). Furthermore, previous research has shown that lower class children do not achieve as well as middle and upper class children (10, 14), in part due to lower motivation (25). The data obtained in the present study corroborate these observations.

The interrelations found between children's perception of teachers' feelings, school achievement, behavior and socio-economic status are particularly significant since the majority of children in the public schools throughout the country come from families of low social class status. It is therefore likely that a lower class child, especially if he is not doing well in school, will have a negative perception of his teachers' feelings toward him. These negative perceptions will in turn tend to lower his efforts to achieve in school and/or increase the probability that he will misbehave. His poor school achievement will aggravate the negative attitudes of his teachers toward him, which in turn will affect his self-confidence, and so on. This vicious entanglement must be interrupted at some point. The point of attack may well be the teacher whose capacity to reflect feelings conducive to the child's growth should be of concern to educators.

Analysis of Variance of Favorability Scores—It was found that the index of favorability was positively related to achievement in school as well as to social class position. It is also evident from this and other studies that achievement in school is correlated with social class position. In order to study the influence of each of these factors on index of favorability, the favorability scores were re-analyzed first, for the three achievement levels within each social class and second, for the three social class groups within each achievement category. The mean favorability indexes for these separate groups are presented in Table V.

Reading Table V vertically, it may be observed that the mean favorability score declines from the upper social class to the lowest social class for each of the achievement categories; this decline is most noticeable between the two highest social class groups and the lowest social class group. It is apparent that the social class variable plays a part in the way a child perceives his teacher's feelings toward him regardless of his achievement in school. Similarly, reading Table V horizontally, the mean favorability score is observed to decrease from the highest achievement level to the lowest within each social class group. The evidence here suggests that achievement in school colors the child's perception of his teacher's feelings, regardless of his social class position. Analysis of variance of the data yielded two significant F ratios. These results indicate that both the factors of social class position and achievement are operating independently in affecting the way a child will perceive his teacher's feelings toward him.

These findings should arouse the educator for they imply that a teacher's reaction to a child is is not solely influenced by the individuality of the child but also by his social class and achievement characteristics.

Differences Among Teachers—It may be assumed that teachers reflect a variety of feelings toward children, either because of their own personality needs, or because of the way they use punishment or praise or for any other reason. These differences from teacher to teacher should be observable in the perceptions of the children affected by them. Table VI presents the mean favorability indexes for the 10 teachers in this study.

It may be observed that the range in mean favorability score is from 2.25 to 2.70. Although the children generally perceived their teachers' feelings more favorably than otherwise, and the actual differences among the classrooms were not large, there were 3 or 4 classrooms with markedly low mean scores. The overall F ratio of 2.95 is significant at less than the .01 level. It should be remembered, at this point, that the classes were selected for better than average ability in reading, which makes the finding of significant differences even more compelling. Teachers do seem to vary in their inclination and/or their capacity to communicate favorable feelings. It seems urgent that teachers be helped to recognize the significance of the feelings which they express toward children, consciously or unconsciously. Some teachers, in addition, may need the help which can only come through a process of self-understanding, in order to avoid or to minimize the expression of negatively-toned feelings toward children, because of their sex, their socio-economic status, their behavior or achievement in school.

Possible Uses of the Checklist—The Checklist of Trait Names, in addition to its use as a research tool, may be adapted to practical school situations. Conceivably, it can be employed for the purpose of teacher selection and guidance. For instance, a principal might wish to select a teacher for a class comprised of underprivileged or trouble-

TABLE V

MEAN INDEXES OF FAVORABILITY FOR THE THREE ACHIEVEMENT CATEGORIES AND FOR THE THREE SOCIAL CLASS GROUPS

	Achievement Category		
	Very Well	Adequately	Below Average
Upper Social Class	2.71* (N=23)	2.61 (N=34)	2.51 (N= 6)
Middle Social Class	2.71 (N=18)	2.60 (N=25)	2.44 (N=14)
Lower Social Class	2.59 (N=12)	2.52 (N=52)	2.34 (N=19)

*The higher the score, the more favorable the child's perception of his teacher's feelings toward him.

TABLE VI

THE INDEX OF FAVORABILITY FOR THE TEN CLASSROOMS

Class	N	Mean	S.D.
4-1	26	2.61*	.26
5-1	18	2.25	.21
5-2	8	2.45	.29
5-3	27	2.62	.17
5-4	23	2.45	.23
5-5	25	2.62	.22
6-1	17	2.57	.08
6-2	20	2.64	.23
6-3	23	2.64	.19
6-4	16	2.70	.10

*The higher the score, the more favorable the child's perception of his teacher's feelings toward him.

some children who are very much in need of acceptance and approval. A good candidate for such a class would be a teacher who can easily project positive feelings. Supervisors of student teachers may find the checklist useful in evaluating the quality of teacher-student relations.

Teachers who are found to communicate largely negative feelings may be advised to participate in some kind of counseling or therapy. Similarly, children whose perceptions are primarily negative and/or distorted can be identified for personality diagnosis and thus be helped in self-understanding or in obtaining a more accurate perception of reality.

Summary

The purpose of the study was to relate children's perception of their teachers' feelings toward them to self-perception, academic achievement, and classroom behavior. A Checklist of Trait Names, consisting of 35 descriptive terms, was administered to 89 boys and 114 girls in grades 4, 5, and 6 in a New York City public school. The children were rated by their teachers for achievement and on a number of behavioral characteristics.

The major findings were: 1) The children's perception of their teachers' feelings toward them correlated positively and significantly with self-perception. The child with the more favorable self image was the one who more likely than not perceived his teacher's feelings toward him more favorably. 2) The more positive the children's perception of their teachers' feelings, the better was their academic achievement and the more desirable their classroom behavior as rated by the teachers. 3) Further, children in the upper and middle social class groups perceived their teachers' feelings toward them more favorably than did the children in the lower social class group. 4) Social class position was also found to be positively related with achievement in school. 5) However, even when the favorability index data were re-analyzed separately for each social class and for each achievement category, the mean favorability index declined with decline in achievement level, regardless of social class position and, similarly, the mean favorability index declined with social class regardless of achievement level. 6) Girls generally perceived their teachers' feelings more favorably than did the boys. 7) Finally, there were some significant classroom differences in the favorability of the children's perception of their teachers' feelings. These findings must be considered in light of the non-random selection of the sample. Nevertheless, it is reasonable to assume that these subjects are representative of the population of New York City elementary school children at these grade levels.

Possible uses of the Checklist were suggested. As a result of this investigation, a number of changes in the Checklist were indicated. The new form along with instructions, used in a study at the junior high school level, is given below.

CHECKLIST (New Form)

We are asking your cooperation in a study being conducted at ____________ Our interest is in the way you think your ____________ teacher feels toward you.

Now read the following and do the example:

On the next page are pairs of words. In each pair, one is the opposite of the other. There are five steps between the pairs of words, as shown below.

PLEASANT___A___.___B___.___C___.___D___.___E___UNPLEASANT

Consider the words PLEASANT and UNPLEASANT. Here is what you are supposed to do. If you feel
your teacher judges you as being pleasant most of the time, put an X in the "A" box. If you feel your
teacher considers you to be unpleasant most of the time, place an X in the "E" box. If you feel your
teacher considers you to be pleasant sometimes and unpleasant sometimes, mark the "C" box. If you
feel your teacher considers you to be somewhat more pleasant than unpleasant, put an X in the "B" box.
If you feel your teacher considers you to be somewhat more unpleasant than pleasant, place an X in the
box marked "D".

Now do the example. Be sure to place an X in the middle of the box which describes most nearly how your teacher feels about you. Do not worry or puzzle over any one word. It is your immediate feeling that we want. Do not omit any word.

Be as honest as you can. Neither your teacher nor your principal will see your paper.

If you do not understand a word, please raise your hand, and the Examiner will come to your seat and explain it to you.

TURN THE PAGE AND BEGIN.

DAVIDSON - LANG

Name ____________________

____________________ ____________________

	A	B	C	D	E			A	B	C	D	E	
Selfish						Unselfish	Wise						Foolish
Obedient						Disobedient	Rude						Polite
Intelligent						Unintelligent	Alert						Lazy
Good						Bad	Passive						Active
Sad						Happy	Reliable						Unreliable
Slow						Fast	Nice						Awful
Clean						Dirty	Unpopular						Popular
Strong						Weak	Curious						Indifferent
Cowardly						Brave	Careless						Careful
Honest						Dishonest	Daring						Afraid
Calm						Nervous	Childish						Mature
Unfair						Fair	Attentive						Inattentive
Graceful						Awkward	Disorderly						Orderly
Kind						Cruel	Ungrateful						Grateful
Unfriendly						Friendly	Respectful						Disrespectful

BIBLIOGRAPHY

1. Allport, G., and Odbert, H. "Trait Names: A Psycho-lexical Study," *Psychological Monographs*, XLVII (1936).
2. Ausubel, D. P., and others. "Perceived Parent Attitudes as Determinants of Children's Ego Structure," *Child Development* XXV (1954), pp. 173-83.
3. Barr, A. S., and Jones, R. E. "The Measurement and Prediction of Teacher Efficiency," *Review of Educational Research*, XXVIII (1958), pp. 256-64.
4. Bills, R. E., and others. "An Index of Adjustment and Values," *Journal of Consulting Psychology*, XV (1951), pp. 257-61.
5. Cook, W. W., "Significant Factors in Teachers' Classroom Attitudes," *Journal of Education*, VII (1956), pp. 274-79.
6. Davis, A. *Social Class Influences Upon Learning* (Cambridge, Mass.: Harvard University Press, 1952).
7. de Groat, A. F., and Thompson, G. G. "A Study of the Distribution of Teacher Approval and Disapproval Among Sixth-Grade Pupils," *Journal of Experimental Education*, XVIII (1949), pp. 57-75.
8. Dixon, N. R. "Social Class and Education," *Harvard Educational Review*, XXIII (1953), pp. 330-38.
9. Dymond, Rosalind F. "A Scale for Measurement of Empathic Ability," *Journal of Consulting Psychology*, XIII (1949), pp. 127-33.
10. Friedhoff, W. H. "Relationships Among Various Measures of Socio-Economic Status, Social Class Identification, Intelligence, and School Achievement," *Dissertation Abstract*, XV (1955), p. 2098.
11. Gage, N. L., and others. "Teachers' Understanding of Their Pupils and Pupils' Ratings of Their Teachers," *Psychological Monographs*, LXIX (1955).
12. Gough, H. G. *Reference Handbook for the Gough Adjective Check List*, Mimeo-

graphed (Berkeley, California: University of California, Institute of Personality Assessment and Research, 1955).

13. Hartshorne, H., and May, H. A. Studies in the Nature of Character, III: Studies in the Organization of Character (New York: Macmillan Co., 1930).
14. Heimann, R. A., and Schenk, Q. F. Relations of Social Class and Sex Differences to High School Achievement," School Review, LXII (1954), pp. 213-21.
15. Jersild, A. T. In Search of Self (New York: Bureau of Publications, Teachers College, Columbia University, 1952).
16. Jourard, S. M., and Remy, R. M. "Perceived Parental Attitudes, the Self, and Security," Journal of Consulting Psychology, XIX (1955), pp. 364-66.
17. Martire, J. G. "Relationship Between the Self Concept and Differences in the Strength and Generality of Achievement Motivation," Journal of Personality, XXIV (1956), pp. 364-75.
18. Merrill, R. M., and Heathers, L. B. "The Use of an Adjective Checklist as a Measure of Adjustment," Journal of Consulting Psychology, I (1954), pp. 137-43.
19. Meyer, W. J., and Thompson, G. G. "Sex Differences in the Distribution of Teacher Approval and Disapproval Among Sixth-Grade Children," Journal of Educational Psychology, XLVII (1956), pp. 285-396.
20. National Education Association, Research Division. "Teacher Opinion on Pupil Behavior," Research Bulletin, XXXIV (1956), pp. 51-107.
21. Perkins, H. V. "Factors Influencing Change in Children's Self Concepts," Child Development, XXIX (1958), pp. 221-30.
22. Perkins, H. V. "Teachers' and Peers' Perceptions of Children's Self Concepts," Child Development, XXIX (1958), pp. 203-220.
23. Reeder, T. A. "A Study of Some Relationships Between Level of Self Concept, Academic Achievement, and Classroom Adjustment," Dissertation Abstract, XV (1955), p. 2472.
24. Rogers, C. R. Client-Centered Therapy (Boston: Houghton-Mifflin, 1951).
25. Rosen, B. C. "The Achievement Syndrome: A Psychocultural Dimension of Social Stratification," American Sociological Review, XXI (1956), pp. 203-11.
26. Sarbin, T. R., and Rosenberg, B. C. "Contributions to Role-Taking Theory: IV. A Method for Obtaining a Qualitative Estimate of the Self," Journal of Social Psychology, XLII (1955), pp. 71-81.
27. Snygg, D., and Combs, A. W. Individual Behavior (New York: Harper and Brothers, 1949).
28. Steiner, I. D. "Self-Perception and Goal-Setting Behavior," Journal of Personality XXV (1957), pp. 344-55.
29. Stevens, P. H. "An Investigation of the Relationship Between Certain Aspects of Self-Concept Behavior and Students' Academic Achievement," Dissertation Abstract, XVI (1956), pp. 2531-2532.
30. Sullivan, H. S. Conceptions of Modern Psychiatry (Washington, D. C.: W. A. White Psychiatric Foundation, 1947).
31. Symonds, P. M. "Characteristics of the Effective Teacher Based on Pupil Evaluation," Journal of Experimental Education, XXIII (1955), pp. 289-310.
32. Terman, L. M., and Tyler, L. E. "Psychological Sex Differences," in L. Carmichael Manual of Child Psychology, Second Edition (New York: Wiley and Sons, 1954).
33. Thorndike, E. L., and Lorge, I. The Teacher's Word Book of 30, 000 Words (New York: Bureau of Publications, Teachers College, Columbia University, 1944).

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 2, December 1960)

PSYCHOLOGICAL SERVICES IN TWENTY-EIGHT ELEMENTARY SCHOOLS OF COLUMBUS, OHIO

OPAL E. PARMER
Board of Education, Columbus, Ohio

THE VALUE of psychological services for a large city school system has never seriously been questioned. The demand usually exceeds the means. This is true not only in a large city, but also in county and state services. This study was done in an effort to clarify the major areas of demand for psychological services in the elementary schools of Columbus, Ohio.

The purpose of this study was: 1) to determine if psychological services were being distributed evenly among the various elementary schools according to their enrollments; 2) to determine if the schools with high achievement records required the same type and amount of services as those with low achievement records; 3) to determine the kinds of difficulties the elementary children were having; and 4) to see if the difficulties were being corrected through parental guidance and other means.

Enrollment

There were twenty-eight elementary schools used in this study. This equalled 30.43 percent (approximately one-third) of the total elementary schools, or an enrollment as of May, 1960, of 14,548 children in grades kindergarten through six. Table I shows the enrollments of the various schools. All but three schools had more boys than girls. There was 8.88 percent more boys than girls in the schools where this study was made. This may be compared with the 5.83 percent more boys than girls in the enrollment of all elementary schools in this study. Data in Table II indicates the number and percent who received psychological assistance acoording to schools: the percent who were boys or girls. In all schools the boys presented more concerns than the girls. This was equally true of the schools having a high or a low achievement record.

Table III compares the number and percent of children who were over-age for the grade when referred to the psychologist for assistance. These children (61.15% of those tested) had been recognized classroom concerns for more than one year. The majority of them had repeated at least one grade. A few had come from other school systems or other states where the attendance laws vary.

A comparison between the high and low achievement schools indicated 49 overage children received assistance from the top quartile group while 110 overage children received assistance from the lowest quartile group.

Achievement

The rank of academic achievement for the schools used in this study was determined by the school's standing when compared with the total elementary schools of the city. The city-wide achievement test scores for sixth grade were used as a basis of comparison.

It was learned that of the twenty-eight schools used, 21.43 percent ranked high, 3.57 percent ranked above average, 28.57 percent ranked average, 17.86 percent ranked below average, and 28.57 percent ranked low.

A comparison between the achievement scores and the socio-economic status of the children within the schools correlated proportionately. The schools are grouped in the Tables according to their academic achievement records for the past two years.

There were 32.14 percent of the schools in this study who had a special class program to assist the children who, on individual intelligence test scores, placed between the 50 and 75 I.Q. range. An additional 7.14 percent of the schools not having a special class program this year plan to include it in their curriculum next year.

One school in this study was designed for handicapped children. The necessity for its having additional psychological services is greater. Yet the overall services used are not as great as would be needed if these children were instructed in their neighborhood school.

Kinds of Difficulties

The reasons for desiring psychological ser-

TABLE I

NUMBERS AND PERCENTAGES OF BOYS MORE THAN GIRLS IN TWENTY-EIGHT SCHOOLS OF COLUMBUS, OHIO, IN MAY, 1960

Schools	Enrollment as of May, 1960	Boys	Girls	Boys more than Girls	Percent more Boys
A	849	443	406	37	8.35
B	551	269	282	-13	-4.60
C	273	144	129	15	10.41
D	276	139	137	2	1.43
E	678	344	334	10	2.93
F	659	350	309	41	11.71
G	463	256	207	49	19.14
H	576	304	272	32	10.52
I	970	497	473	24	4.83
J	243	134	109	25	18.65
K	606	316	290	26	8.22
L	184	106	78	28	26.41
M	281	152	129	23	15.13
N	343	185	158	27	14.59
O	800	418	382	36	8.60
P	365	204	161	43	21.07
Q	616	320	296	24	7.50
R	297	161	136	25	15.52
S	475	269	206	63	23.42
T	682	367	315	52	14.16
U	538	290	248	42	14.48
V	294	144	150	- 6	-4.0
W	819	415	404	11	2.65
X	697	372	325	47	12.74
Y	497	229	268	-39	-14.55
Z	536	278	258	20	7.19
AA	549	288	261	27	9.37
BB	431	218	213	5	2.29
Totals	14,548	7,612	6,936	676	8.88
All Elementary of Columbus	531,035	27,315	25,720	1,595	5.83

TABLE II

NUMBERS AND PERCENTAGES OF CHILDREN WHO RECEIVED INDIVIDUAL PSYCHOLOGICAL ASSISTANCE IN TWENTY-EIGHT ELEMENTARY SCHOOLS OF COLUMBUS, OHIO, 1959-1960

School	Number tested	Percent of enrollment	Percent who were boys	Percent who were girls
A	20	2.35	85.00	15.00
B	11	1.99	72.73	27.27
C	7	2.56	57.15	42.85
D	8	2.90	75.00	25.00
E	23	3.39	65.22	34.78
F	33	5.00	66.67	33.33
G	9	1.08	100.00	0.0
H	12	2.08	100.00	0.0
I	27	2.78	70.38	29.62
J	6	2.46	100.00	0.0
K	13	2.14	76.93	23.07
L	28	14.67	67.86	32.14
M	1	.35	100.00	0.0
N	9	2.62	66.67	33.33
O	23	2.87	52.18	47.82
P	13	3.56	84.62	15.38
Q	16	2.59	68.75	31.25
R	5	1.68	60.00	40.00
S	23	4.86	78.27	21.73
T	41	6.01	65.86	34.14
U	7	1.30	57.15	42.85
V	14	4.76	71.43	28.57
W	32	3.90	65.63	34.37
X	23	3.29	60.87	39.13
Y	15	3.02	60.00	40.00
Z	27	5.03	92.60	7.40
AA	14	2.55	78.58	21.42
BB	13	3.01	76.93	23.07
Office	47		53.20	46.80
Total	520			

TABLE III

NUMBERS AND PERCENTAGES OF CHILDREN WHO WERE OVER AGE FOR THE GRADE WHEN TESTED IN THE TWENTY-EIGHT ELEMENTARY SCHOOLS OF COLUMBUS

School	Number tested	Number over age	Percent who were over age
A	20	12	60.00
B	11	1	9.99
C	7	2	28.57
D	8	5	62.50
E	23	9	38.69
F	33	15	45.45
G	9	5	55.55
H	12	4	33.33
I	27	14	51.82
J	6	5	83.84
K	13	6	46.15
L	28	21	75.00
M	1	1	100.00
N	9	4	44.44
O	23	11	47.82
P	13	10	76.92
Q	16	12	75.00
R	5	3	60.00
S	23	7	30.43
T	41	23	56.09
U	7	5	71.42
V	14	13	92.85
W	32	24	75.00
X	23	21	91.30
Y	15	7	46.66
Z	27	26	96.29
AA	14	8	57.14
BB	13	11	84.61
Office	47	33	70.21
Total	520	318	61.15

vices have been arranged into five areas. Children were classified according to their area of greatest need. They were referred by the teacher through the principal to the psychologist for assistance.

Testing for ability was done when the school needed to know how much the child was capable of learning at that time, or how much he could absorb intellectually.

Requests because of behavior were for those children who, because of their mode of acting in the presence of others, or toward them, created a situation in which structured learning was impaired.

Referrals because of social adjustment were for those children who had been unsuccessful in establishing an acceptable relationship between themselves and their environment.

Suspected medical reasons were those where the physical aspects of the child were to be considered. Such children were often suspected of having a psychosis or brain injury. The medical cases were tested and rechanneled through the school health department if this seemed advisable.

Special class children are retested by an individual psychological examination every third year in compliance with Ohio School Law. The special class is structured for instruction at the child's mental level of development. Only children having an I.Q. score of from 50 to 75 on an individual psychological examination and who the school thinks would profit from this type homogeneous instruction are placed in these classes. The classes are housed within the local community's school.

Outside agencies occasionally requested our assistance in helping them understand some of the children's needs. A written report of our findings was sent to them.

Table IV indicates the reasons why individual psychological testing and assistance were given to the 520 children. The greatest need was in the area of ability testing. In fairness to the schools it must be admitted that many children were referred for more than one reason. This was especially true of those in the Ability Group.

A comparison between the high and low achievement schools indicates 54 tests were needed for children from the top quartile group while 75 tests were needed for determining ability among the lowest quartile group.

City-wide group testing to determine ability was given in the second and fifth grades of the elementary schools.

Social adjustment was the next largest area. The behavior and adjustment areas were so closely related that most children who had not adjusted environmentally were also having school behavior difficulties. This was reversely true of the children classified in the behavior section. No child was classified in more than one area, however.

A comparison between the high and low achievement schools indicates nine tests were used for children from the top quartile group who were having behavior difficulties. Seventeen children were having major behavior difficulties among the lowest quartile group.

Similarly, the top quartile group required 37 tests because of adjustment and the lowest quartile group required 32.

Children of the top quartile group surpassed the children of the lowest quartile group 11 to 4 in medical needs.

Outside agencies requested assistance for 24 children. These children are placed in one of the other classifications according to their primary difficulty. There were three tests given in the child's home because of the child's physical impairment. Four tests were given at the request of private physicians, seven were for the Juvenile Court; the others were for other agencies of the city such as the Catholic Guidance Center or Children's Hospital.

Tables V and VI show the grade level by number and by percent of the children receiving psychological assistance. Testing at any grade level was not done if other assistance and the child's group test scores could be used in solving the difficulty. Kindergarten test scores were recognized as unstable, especially if the child was hyperactive. Most of the kindergarten testing was needed for the top quartile schools.

Testing to identify the gifted was not done since the group intelligence tests could be used for this purpose at present and other difficulties seemed more acute.

The largest number of tests was given at Grade 1. The demands at sixth grade were next. A comparison between the high and low achievement schools indicated more testing (35 tests) was needed for the top quartile schools in Grade 1, while more testing (31 tests) was needed for the lowest quartile schools in Grade 6. Most testing and assistance for the top quartile schools was done in the primary grades while most assistance was given the lowest quartile schools in the intermediate grades (grades 4, 5, and 6).

Table VII shows the intelligence quotient range by number and by percent for the 520 children given psychological testing and assistance in the twenty-eight elementary schools of this study. The cut-off points for the intelligence quotient figures were obtained from the suggested range in the Stanford-Binet Intelligence Test manual. Children of average intelligence (90 - 110 I. Q.) were most frequently referred for assistance. The children who were classified as having below average intelligence also seemed to have much difficulty in school.

A comparison between the high and low achievement schools indicated a need for 52 tests

TABLE IV

REASONS WHY INDIVIDUAL PSYCHOLOGICAL TESTING AND ASSISTANCE WERE GIVEN IN TWENTY-EIGHT ELEMENTARY SCHOOLS OF COLUMBUS, OHIO, 1959-1960

School	Ability		Behavior		Adjustment		Medical		Special Class Retests		Outside Agencies*	
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
A	8	40.00	3	15.00	7	35.00	2	10.00				
B	4	36.40			6	54.60	1	9.00				
C	3	42.85	2	28.57	1	14.29	1	14.29			1	14.29
D	6	75.00			1	12.50	1	12.50				
E	8	34.79	1	4.34	13	56.53	1	4.34				
F	20	60.60	2	6.06	6	18.19	5	15.15				
G	5	55.55	1	11.11	3	33.34						
H	4	33.33	2	16.67	3	25.00	2	16.67	1	8.33		
I	18	66.60			7	26.00	2	7.40				
J	4	66.66			2	33.34						
K	5	38.47	4	30.75	3	23.09	1	7.69				
L	26	92.86					1	3.57	1	3.57		
M					1	100.00						
N	5	55.55			4	44.45						

O	16	69.58			6	26.08	1	4.34				
P	4	30.77	2	15.39	7	53.84						
Q	8	50.00	2	12.50	5	31.25	1	6.25				
R			2	40.00	3	60.00					1	20.00
S	8	34.78	1	4.35	12	47.83	2	8.70				
T	23	56.10	10	24.39	8	19.51					2	4.88
U	2	28.57	2	28.57	3	42.86					2	28.57
V	6	50.00			5	28.57			3	21.43	1	7.14
W	17	53.10	3	9.37	9	28.17	1	3.12	2	6.24	2	6.24
X	6	26.10	4	17.40	7	30.42	1	4.35	5	21.73	1	4.34
Y	13	86.64	1	6.68			1	6.68				
Z	21	77.80	3	11.10	2	7.40	1	3.70			3	11.10
AA	4	28.58	5	35.71	5	35.71						
BB	8	61.53	1	7.69	4	30.78					3	23.06
Office	10	21.28	3	6.38	10	21.28	5	10.64	19	40.42	8	17.02
Total	262		54		143		30		31		24	

---

* Children classified under Outside Agencies are also placed in one of the other classifications depending on their primary difficulty.

TABLE V

GRADE LEVEL BY NUMBER OF CHILDREN GIVEN PSYCHOLOGICAL TESTING AND ASSISTANCE IN TWENTY-EIGHT ELEMENTARY SCHOOLS OF COLUMBUS, OHIO

School	Special	Kindergarten	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6
A		1	7	1	1	5	4	1
B			2	2	3		1	3
C			4				2	1
D			2	2		1	3	
E		3	4	1	4	3	5	3
F		1	15	5	3	6		3
G			1	3	1		3	1
H		1	5	2	1	2	1	
I		3	6	11	4	1		2
J			1					5
K			6	2	3	1	1	
L	1		14		2	1	5	5
M					1			
N			1		1			7
O		2	4			3	2	12
P		2		2	2	2	4	3
Q			2	1	3	3	2	5
R			2			1	1	1
S		5	8	3	4	2	1	
T		3	5	10	6	1	2	14
U				3		3		1
V	3			2	1	3		5
W	2		10	3	4	4	5	4
X	5		1	5	6	2	2	2
Y			1	1	3	5	2	3
Z			1	8	4	1	6	7
AA			2	1	1	4	2	4
BB				4	1	1	1	6
Office	19	2	12	6	2	3	3	—
Total	30	21	118	76	61	58	58	98

TABLE VI

GRADE LEVEL BY PERCENT OF CHILDREN GIVEN PSYCHOLOGICAL TESTING AND ASSISTANCE IN TWENTY-EIGHT ELEMENTARY SCHOOLS OF COLUMBUS, OHIO

School	Special	Kindergarten	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6
A		5.00	35.00	5.00	5.00	25.00	20.00	5.00
B		18.18	18.18	27.27			9.10	27.27
C			57.14				28.57	14.29
D			25.00	25.00		12.50	37.50	
E		13.04	17.39	4.35	17.39	13.04	21.75	13.04
F		3.03	45.46	15.15	9.09	18.18		9.09
G			11.11	33.34	11.11		33.33	11.11
H		8.33	41.67	16.67	8.33	16.67	8.33	
I		11.10	22.20	40.80	14.80	3.70		7.40
J			16.67					83.33
K			46.15	15.38	23.09	7.69	7.69	
L	3.57		50.00		7.14	3.57	17.86	17.86
M					100.00			
N			11.11		11.11			77.79
O		8.70	17.39			13.04	8.70	52.17
P			15.39		15.39	15.39	30.77	23.06
Q			12.50	6.25	18.75	18.75	12.50	31.25
R			40.00			20.00	20.00	20.00
S		21.74	34.78	13.04	17.39	8.70	4.35	
T		7.32	12.19	24.39	14.63	2.44	4.88	34.15
U				42.86		42.86		14.28
V	21.43			14.28	7.14	21.43		35.72
W	6.25		31.25	9.37	12.50	12.50	15.62	12.50
X	21.73		4.34	21.73	26.08	8.69	8.69	8.69
Y			6.68	6.68	20.00	33.33	13.33	20.00
Z			3.70	29.65	14.81	3.70	22.22	25.92
AA			14.28	7.14	7.14	28.58	14.28	28.58
BB			30.78	7.69	7.69		7.69	46.15
Office	40.42	4.26	25.53	12.77	4.26	6.38	6.38	

TABLE VII

INTELLIGENCE QUOTIENT RANGE BY NUMBER AND BY PERCENT FOR CHILDREN GIVEN PSYCHOLOGICAL TESTING AND ASSISTANCE IN TWENTY-EIGHT ELEMENTARY SCHOOLS

School	Below 50		50-75 IQ		76-89 IQ		90-110 IQ		111-125 IQ		126+ IQ		Total
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	
A			4	25.00	3	15.00	9.	40.00	3	15.00	1	5.00	20
B					1	9.00	4	36.40	5	45.60	1	9.00	11
C					3	42.85	4	57.15					7
D			2	25.00	2	25.00	3	37.50	1	12.50			8
E			3	13.04	7	30.45	8	34.78	2	8.69	3	13.04	23
F			5	15.15	7	21.21	19	57.38	2	6.06			33
G			1	11.11			5	44.44	1	11.11	2	22.22	9
H			3	25.00	1	8.33	7	58.34	1	8.33			12
I			9	33.40	10	37.00	8	29.60					27
J			2	33.34	1	16.67	3	49.99					6
K			2	15.38	2	15.38	7	53.86	1	7.69	1	7.69	13
L	1	3.57	9	32.14	9	32.14	7	25.01	1	3.57	1	3.57	28
M									1	100.00			1
N			1	11.11	3	33.32	3	33.32	1	11.11	1	11.11	9
O			4	17.39	7	30.45	8	34.78	3	13.04	1	4.34	23
P	1	7.70	4	30.77	4	30.77	3	23.06	1	7.70			13
Q			4	25.00	6	37.50	5	31.25	1	6.25			16
R			2	40.00			2	40.00	1	20.00			5
S			1	4.35	11	47.83	8	34.78	3	13.04			23
T			10	24.39	13	31.71	15	36.58	1	2.44	2	4.88	41
U			2	28.57	1	14.28	4	57.15					7
V			8	57.16	4	28.56			2	14.28			14
W	4	12.50	20	62.50	4	12.50	3	9.37	1	3.13			32
X			11	47.86	8	34.77	3	13.03	1	4.34			23
Y			7	46.67	5	33.33	3	20.00					15
Z			11	40.76	13	48.14	3	11.10					27
AA			2	14.28	5	35.71	6	42.87	1	7.14			14
BB			7	53.84	4	30.78	1	7.69	1	7.69			13
Office	16	34.04	8	17.02	11	23.40	10	21.28	2	4.26			47
Total	22		142		145		161		37		13		520

TABLE VIII

OUT-OF-SCHOOL PARENT CONFERENCES BY NUMBER AND PERCENT FOR CHILDREN GIVEN PSYCHOLOGICAL TESTING AND ASSISTANCE IN TWENTY-EIGHT ELEMENTARY SCHOOLS OF COLUMBUS, OHIO

School	Number Tested	Number of Parent Conferences	Percent of Parent Conferences
A	20	5	25.00
B	11	7	63.63
C	7	2	28.57
D	8	0	0.0
E	23	12	52.17
F	33	3	9.09
G	9	3	33.34
H	12	5	8.33
I	27	1	3.70
J	6	0	0.0
K	13	1	7.69
L	28	5	17.86
M	1	1	100.00
N	9	1	11.11
O	23	3	13.04
P	13	6	46.12
Q	16	1	6.25
R	5	3	60.00
S	23	6	26.08
T	41	11	26.82
U	7	2	28.57
V	14	2	14.28
W	32	2	3.12
X	23	1	4.35
Y	15	3	20.00
Z	27	0	0.0
AA	14	6	50.00
BB	13	0	0.0
Office	47	37	78.72
Total	520	129	

TABLE IX

REQUESTED BUT UNABLE TO TEST BECAUSE OF LACK OF TIME

School	Ability	Behavior	Adjustment	Medical	Special Class	Total
A	1			1	1	3
B						0
C	1					1
D						0
E	11		5	2		18
F	1		1			2
G	9	1				10
H	5	2	3			10
I	6		3	1	9	19
J	3					3
K	1					1
L	1			1		2
M			1			1
N	11				1	12
O	1				1	2
P	1	1	1			3
Q						0
R	1	3				4
S	4			2		6
T	9	1				10
U	1					1
V	1			1		2
W					1	1
X	22	2	1	1		26
Y		2			2	4
Z				1		1
AA	2	1				3
BB	3	—	2	—	1	6
Total	95	13	17	10	16	151

from the top quartile group at the average ability level and 19 tests for the lowest quartile group.

Elementary Guidance

The elementary guidance program was primarily a part of the elementary classroom curriculum. There was a consistent effort on the part of school personnel to use the team approach in helping the child solve or adjust to his difficulties. The child's family, teacher, and principal were integral parts of the team. The psychologist frequently participated in a team discussion concerning the child's difficulties. Such conferences were usually held after school so the teacher might be present and the working parent would not need to be absent from his employment for long. The length of the parent conference varied according to the child's difficulties and the understanding of those concerned in being able to chart corrective measures. Conferences were usually an hour in length, occasionally longer. Since this was an after-school responsibility, it was felt a more successful conference could be held in a relaxed, unhurried manner.

Table VIII shows the number of out-of-school parent conferences and compares them as a percent with the number of children in each school given psychological testing and assistance.

A comparison between the high and low achievement schools indicated there were more than double (32) the out-of-school parent conferences concerning children from the top quartile group. There were fourteen for the lowest quartile group.

There were 182 calendar school days for the year 1959-1960. There were 129 out-of-school conferences with parents.

Table IX indicates the major area of concern for children who could not be tested because of lack of time. This 151 referrals compares with the 142 who could not receive psychological assistance during the preceding school year for the same reason. Had all requests received psychological assistance, the area for Ability Testing would have been appreciably greater in Table IV.

There is a similar comparison for each school between the number of children receiving psychological testing and assistance for the 1959-1960 school year and previous school years.

Summary

Boys were more frequently referred for psychological assistance than girls.

An average three and one-half percent of the children in these twenty-eight elementary schools received psychological assistance.

Approximately three-fifths of those receiving assistance were over-age for the grade.

Most children had more than one major area of difficulty before being seen by the psychologist. Twenty-nine percent were continuing difficulties for more than two years duration.

Grade 1 seemed to need most assistance in schools of high academic achievement. Grade 6 needed most services in schools of a low achievement level.

Children of average intelligence were most frequently referred for psychological assistance.

The demand for psychological services increased as the I. Q. range of the children lowered.

More than twice as many out-of-school parent conferences were held for children from the high academic achievement schools as for those from the low achieving schools.

The requests for assistance were greater than the amount of time available for such services, even though the psychologist's days were lengthened beyond those of the average classroom teacher.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 2, December 1960)

GALVANIC SKIN RESPONSES DURING PROBLEM SOLVING

FRANK W. BANGHART
University of Virginia

THE RELATIVELY low correlation (.40 to .60) between intelligence and achievement suggests that systematic research in other areas is appropriate.

This exploratory study was an attempt to investigate the possible relationships between physiological responses (as inferred from GSRs) and certain specified variables, i.e., interest, intelligence, anxiety, academic grades, and grouping.

Background

The role of the autonomic nervous system, and in particular, the activity shown by the changes in galvanic skin resistance, has long been recognized as an integral factor in the emotional functioning of the organism. The ANS, as soon as a stimulus effects a change in the organism, begins a reaction to return the system to homeostasis. Ruff, et al. (22) specify that the continuous patterns of the ANS which are perceived in various instruments show that the ANS is working in conjunction with centers which regulate behavior.

McCleary (17) discusses three theories of the nature of the galvanic skin response: the muscular, vascular and sweat-gland theories. After summarizing studies by proponents of each of these hypotheses, he dismisses the first two and concludes, with Darrow (3) that GSR is somehow connected to sweat-gland function as a presecretory activity. Darrow's early hypothesis was that the increased secretory activity preceding actual sweating, triggered by the sympathetic nervous system, was associated with the fall in skin resistance known as GSR. In this way the skin resistance serves as an index of the activation of the organism, by showing the activity of the sympathetic system.

The pathways used by the GSR are not yet clearly understood. McCleary draws the following conclusion: The sympathetic nervous system and the post-ganglionic fibers seem to be the "final pathway" of the GSR. Whether the GSR pathway crosses the spinal column is still being debated; McCleary reports research (17) which indicates that it does.

The polygraphic technique, known popularly as the "lie detector", has received much publicity, not all of which has been favorable. However, this technique is based upon the familiar hypothesis that changes in emotion, especially suppressed fear, appear in a physiological manner and can be detected by the use of a polygraph. This instrument measures, in addition to the GSR, heart rate, blood pressure, respiration, and, in some cases, muscular tension. For an excellent description of this instrument and the polygraphic technique, the reader is referred to Lee (16).

Most police investigators prefer breathing and blood pressure to the GSR as indices of tension, because they feel that the GSR is too sensitive a measure. However, Jost (8) reports that this is true only when too strong a stimulus is used; under more subtle stimuli the GSR may point up reactions which would be lost on a grosser measurement. Thus, the GSR, if handled properly, can be especially valuable to research, since it is capable of very delicate measurement.

Jost gives an excellent picture of the ANS and polygraphic techniques in research. He comments that since the polygraph measures the activities of the ANS as manifested in heart rate, muscular tension, blood pressure, respiration, and skin resistance, and the ANS plays an important role in the adjustment of the individual to a situation, the polygraphic technique can be used to measure and record the adjustment of the individual to his environment, and environmental stimuli. In this case, the actual degree of fluctuation is not nearly so important as the pattern which is produced.

The relationship between emotion and an increase in activation as shown by a rise in skin conductance, is described by Schlosberg (23). He describes his work in charting an "emotion solid," similar to the Munsell color solid. The items are a series of pictures, arranged on two axes—pleasant-unpleasantness, and fear-disgust.

---

*This study was made possible by a research grant from the Group Psychology Branch, Office of Naval Research, Contract Number NONR 474 (8).

Schlosberg suggests that a third dimension, which would take care of variations which are irreconcilable in a two-dimensional figure, and which would correspond to the axis of brightness in the Munsell solid, could be an axis based upon the level of activation of the picture's subject, as shown by the GSR, from sleep to tension.

Since the GSR gives such an excellent index to the emotional involvement of the individual, much research has been done with various psychological tests and their impact on the organism. Levy (15) took the GSRs of subjects during the presentation of the Rorschach test. By presenting the cards, not in the customary manner, but rotated, she hoped to be able to find if any of the cards had a particular "affective tone" which caused a greater response than the others. In the group of normals which she used, she found that the GSRs to the cards varied with the individual; however, there was no particular card or group of cards which seemed to elicit a greater response from the group as a whole. Hugh et al., (7) ran a similar test to determine if there was a greater response to color than to non-color cards. He could find "no appreciable variation" between responses on the color cards as a group and the non-color cards, as measured by the GSR. In reporting another phase of the same study, Jost and Epstein (9) point out the adjustment which takes place in the test situation. At first the GSRs were high and the period of recovery back to base reading quite long. As the subject became adjusted to the testing situation, his responses and the necessary recovery time decreased. A similar pattern of adaptation is shown in the study by Martin (20) who found that the GSRs of twenty subjects tended to decrease for each session of five successive interviews, showing the carry-over of this adaptation from one period of time to another.

Schwartz (25) reports a study on the effects of the Picture Frustration Study in the Galvanic Skin Resistance. It is interesting to note that certain items which, on the basis of our culture, might theoretically be highly-charged situations (namely, lying and unfaithfulness) receive a very high mean deflection from the twenty students taking the test. Since this test is one in which the subject is asked to project himself into the situation, the GSRs indicated here apparently show graphically the degree of involvement in the frustrating situation which the individual is able to achieve.

In an effort to detect malingering, and the reasons therefore, on the MMPI, Cofer et al., (2) record the GSRs for a number of testees. In later interviews the subjects were asked to explain their reactions to items on which they had shown a very large or a very slight deflection. No correlation was found between high GSRs and an item of serious impact to the testee. Contradictory as this may seem in the light of research reported elsewhere in this paper, this study seems to have validity. It may be possible that the GSR serves as a better index to the subject's response to an item on a personality test than does his answer in a subjective interview.

Herr and Kobler (5, 6) report two experiments in which the reactions of normals and neurotics on a word association test were compared with their GSRs. The variations for the neurotic group between the mean GSRs of the words are significantly greater than for the normals. Herr and Kobler feel that by analyzing a person's GSR pattern made during such a test, and noting the ratio of his responses to each other, an index can be made which would distinguish the neurotic from the normal. Paintal (21) reports a study of 450 psychotics and 450 normals. Taking the GSRs of each of these subjects, when subjected to shock and threat of shock, he found that although the physiological mechanism which produced the GSR is not impaired in the psychotic, there was markedly less response in this group than there was in the control group.

In 1949, McCleary and Lazarus (18) reported the use of the GSR to determine the subliminal perception of nonsense syllables. Their study shows that although the subjects could not consciously identify the syllable which was flashed tachistoscopically at a high speed, he was able to identify it subliminally (as shown by his GSR) if it presented a threat of shock to him. In this same area, McGinnies (19) describes a study using emotionally toned and neutral words. The GSR was measured for the short period before the subject recognized and reported the word. For the critical words the GSR was significantly higher. These stimulus words, McGinnies believes, are clues to deeply imbedded anxiety which, as a part of the reaction of the ANS, can be measured by the galvanic skin response. To avoid further anxiety the subject attempts to avoid recognizing the words; this effort, in turn, makes the GSR deflection even more pronounced.

The GSR has been used recently, as an instrument of research at the Aero Medical Laboratory, Wright Air Development Center, Wright-Patterson Air Force Base. Here Levy, Burch et al., (1, 12, 13, 14, 22) have done research on the patterns of the non-specific responses in skin conductance, as a possible clue to monitoring the consciousness of satellite and other space vehicle crews. Although the onset of sleep cannot be immediately determined from the GSR alone, the rise in skin resistance produced by sleep presents a pattern characteristic enough to be readily identifiable, once this state is attained. Since it has been found that the frequency of non-specific responses is inversely correlated with the level of arousal, the various states of arousal can be determined, for a particular individual. Arousal, participation and awareness of the environment are all considered as factors in consciousness; therefore, the GSR can be used to measure the consciousness of the individual.

The pattern of responses from a "signature" of the individual which these investigators consider to be indicative of his personality. Levy et al., (12) describes three types of patterns. One belongs to the hypo-responsive individual. The GSR write-out is flat and relatively low, indicating that this is a person who maintains a high degree of tension at most times. On the other end of the scale is the individual whose record shows a great deal of fluctuation. This is the hyper-responsive individual, who relaxes easily, causing frequent rises in GSR, then brings himself to attention, causing a lowering, as he faces a change of environment. The normo-responsive group show some fluctuations in their GSR record, but they are not so extreme or so frequent as the hypo-responsive group.

Duffy and Lacey demonstrated, in 1946, that the pattern mentioned in connection with the hypo-responsive group is connected, in some cases, with beginning a new task. They found a "saw-toothed pattern" appearing when their subjects were "alerted for a psychophysical task". The initial rise was compensated for as soon as the individual became accustomed to his task.

The Wright-Patterson investigators find that the GSR shows periods of sleep by rise of basal skin resistance, and periods of arousal by a lowering of the resistance. Relaxation and drowsiness are marked by a gradually rising resistance, with fluctuations, as stimuli impinge on the subject.

Therefore, the conclusion might be drawn that the more a subject is involved in his environment, the lower and less variable will be his skin resistance.

Lacey et al., (10) discuss the different types of response patterns possible, considering the various types of autonomic responses. Some people follow a specific pattern of response, whatever the stress; another group may fluctuate greatly from stress to stress; still others use first one pattern of responses, then another. From individual to individual, then, the reactions to a particular stress or situation may vary greatly.

An increase in level of activation, recordable by polygraphic means, may come about from three factors: stimulation from the environment, the occurrence of an idea, or preparation to take some sort of action (23). Landis and Hunt (11) verify this when they say that the GSR is associated with subjective tension—another way of saying "level of activation".

Subjects

The subjects consisted of twenty-two undergraduates who volunteered for the experiment. The general procedure involved the administration of the Taylor Anxiety Scale, Wonderlic Test and the California Occupational Interest Inventory. Immediately following the testing session, the subjects were run through the experimental procedure in an ABAB sequence (i. e., a group (A) non-group (B) basis).

Apparatus

The instrumentation used in these experiments consisted of a Hunter Cardmaster, a Wheatstone bridge, a General Radio type 715-A direct current amplifier, and an Esterline-Angus 5 milliampere graphic recorder. As indicated in Figure 1, a stimulus is provided to the subject by a Hunter Cardmaster model 340. By means of finger electrodes, a battery, and a capacitively coupled Wheatstone bridge, the change of galvanic skin response is detected and then fed into the General Radio DC amplifier which affords sufficient amplification to drive the Esterline-Angus Recorder.

The Hunter Cardmaster was loaded with 37 $3\frac{1}{2} \times 6\frac{1}{4}$ inch plastic stimulus cards. This mechanism, as utilized, enabled the subject to view the stimulus cards individually at any rate or duration which he desired. This control of stimulus was effected by operation of a pushbutton held in the subject's left hand.

Finger electrodes were connected to the right hand on the tips of the first and fourth fingers. Electrical contact was assured by coating the fingers and the electrodes with Sanborn Radux paste and by strapping the electrodes to the fingers with a firm pressure. The electrodes were connected to the Wheatstone bridge, as was the 7. 5 volt battery, and the bridge output was capacitively coupled to the GR DC amplifier through a . 043 microfarad Vitamin Q condenser. This condenser and the input circuit formed a frequency selective filter with a time constant of a half second. Thus only variations of galvanic skin response with a fundamental frequency of 2 cps or greater were recorded. The circuit employed is shown in Figure 2.

The General Radio DC amplifier is a three stage unit having a sensitivity of 100 millivolts input for a full-scale 5 milliamperes output. The unit is stable and calibrated, permitting accurate measurement of dermal resistance at any given time. The high frequency response of the amplifier was found to be considerably above that of the Esterline-Angus graphic recorder tracking capability. Thus changes in GSR above 15 cycles per second were not detected by this system.

Reflecting the net effect of the high frequency response limit and the capacitive coupling, the equipment was used to detect only those variations of GSR within the fundamental range of 1-15 cps. Very gradual changes, or variations whose characteristics were under 1 cps, were not recorded. Thus the measuring apparatus was sensitive only

FIGURE 1

BLOCK DIAGRAM OF APPARATUS

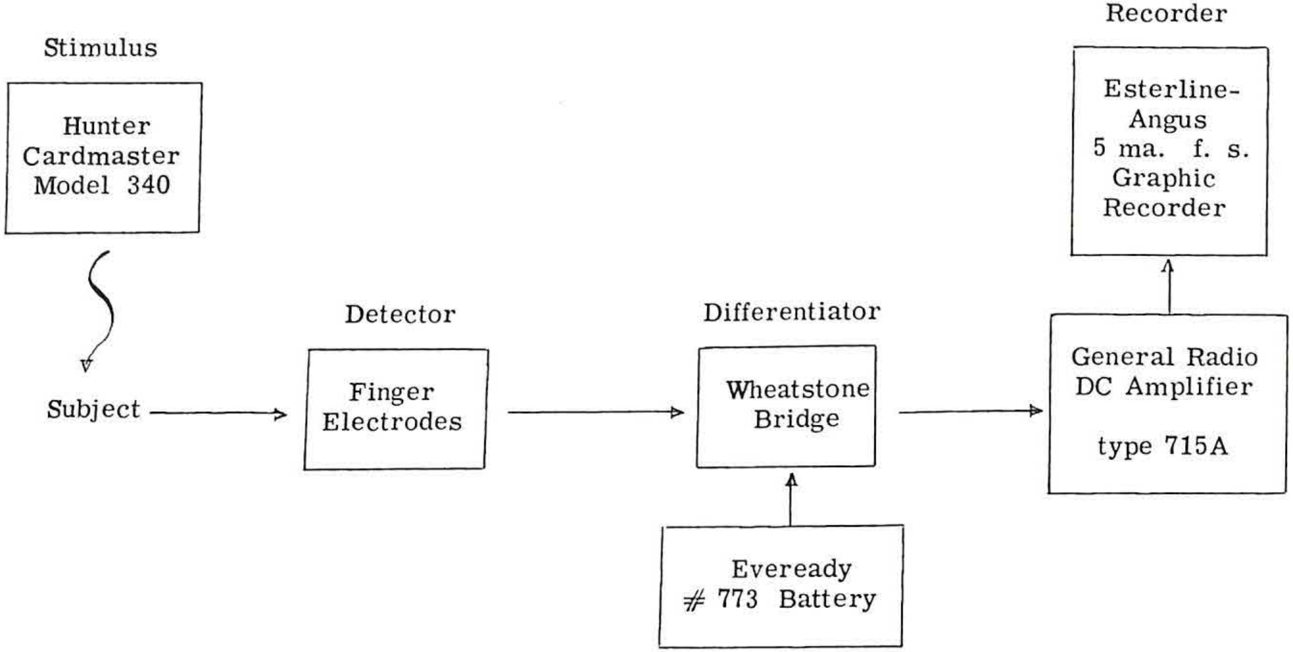

Stimulus
Hunter Cardmaster Model 340
Subject
Detector
Finger Electrodes
Differentiator
Wheatstone Bridge
Eveready # 773 Battery
Recorder
Esterline-Angus 5 ma. f. s. Graphic Recorder
General Radio DC Amplifier type 715A

FIGURE 2

CIRCUIT DIAGRAM OF WHEATSTONE BRIDGE

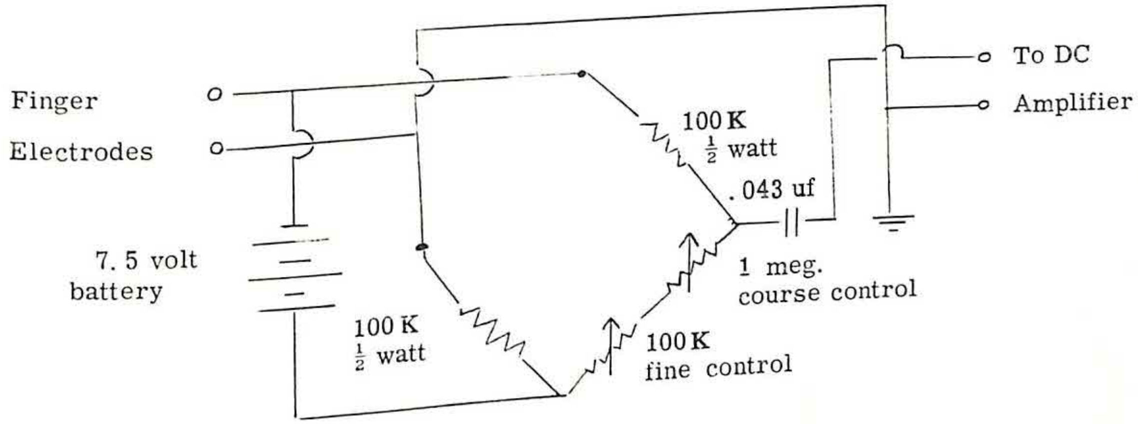


FIGURE 3

GSR AND LEVEL OF INTEREST

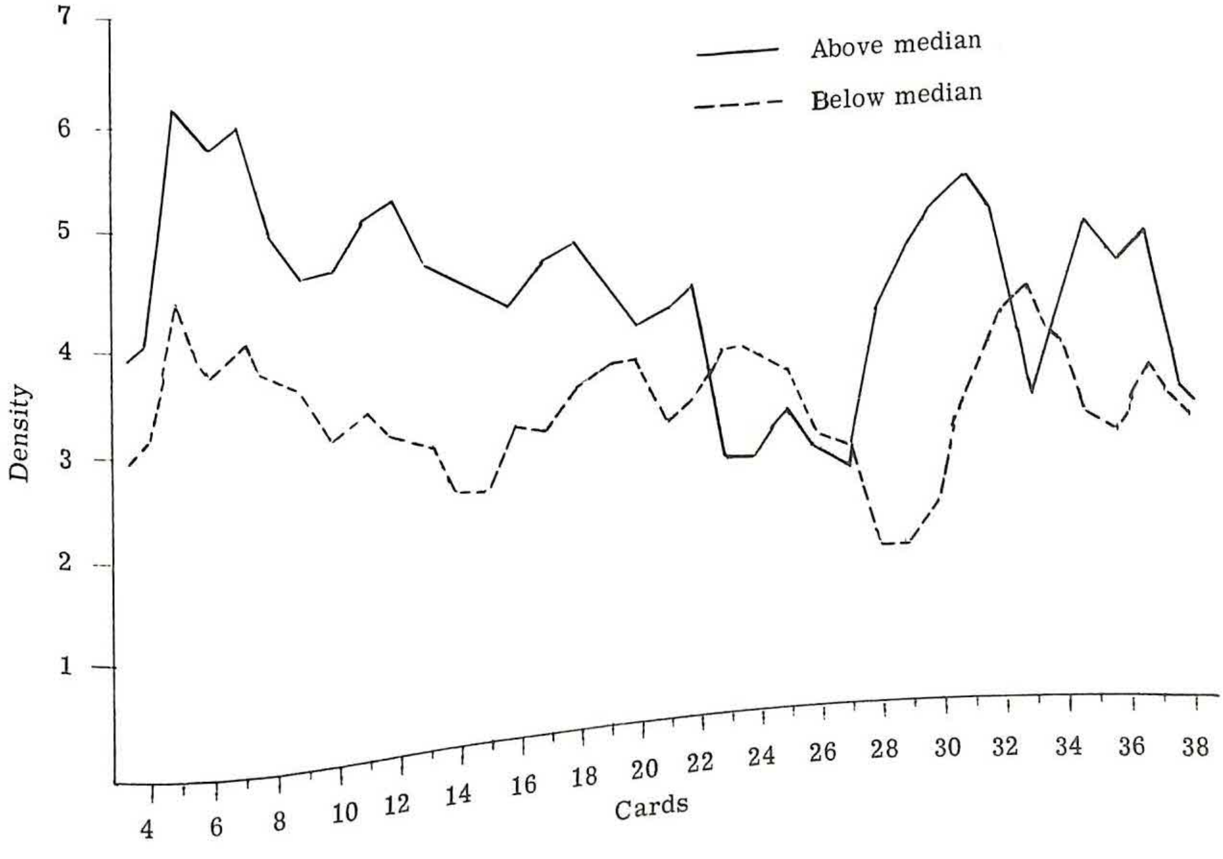


FIGURE 4

GSR AND MENTAL ABILITY

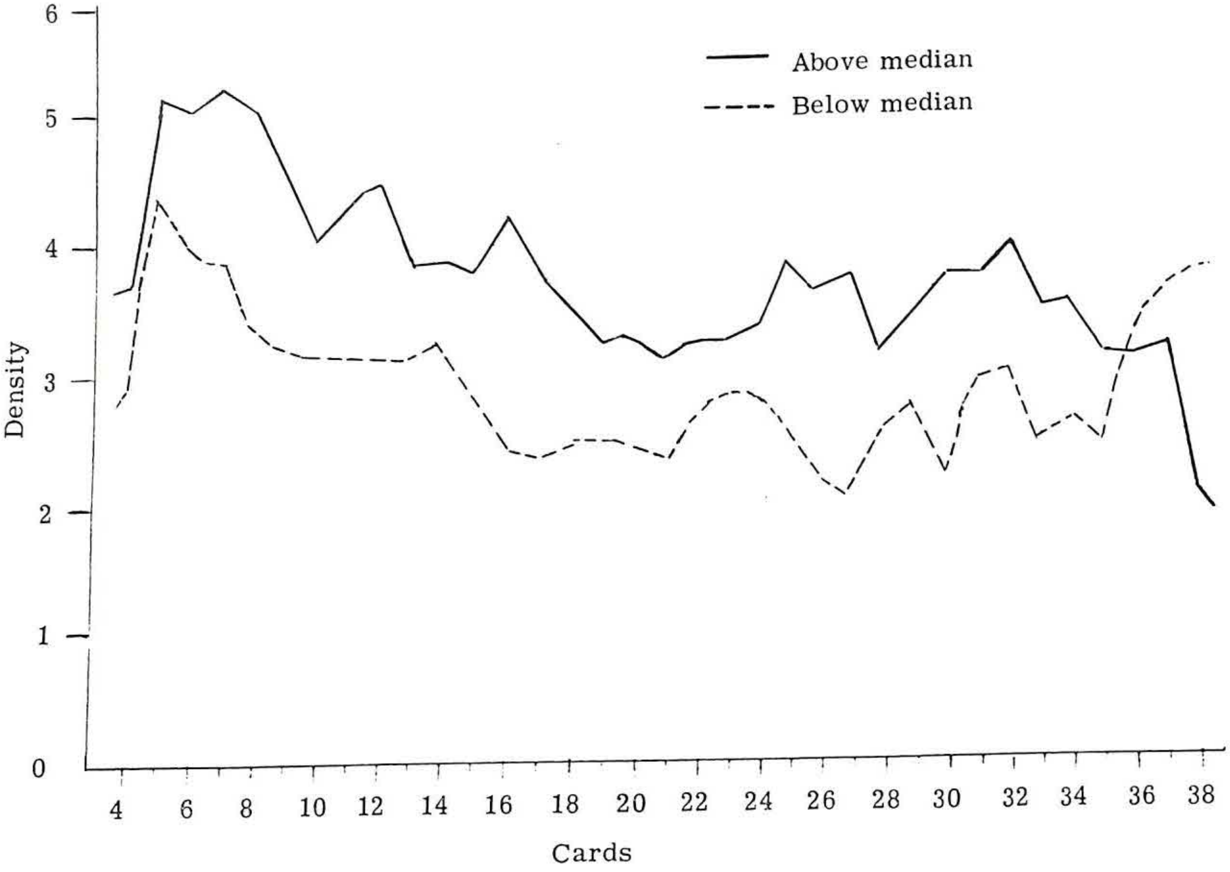


FIGURE 5

GSR AND GROUPING

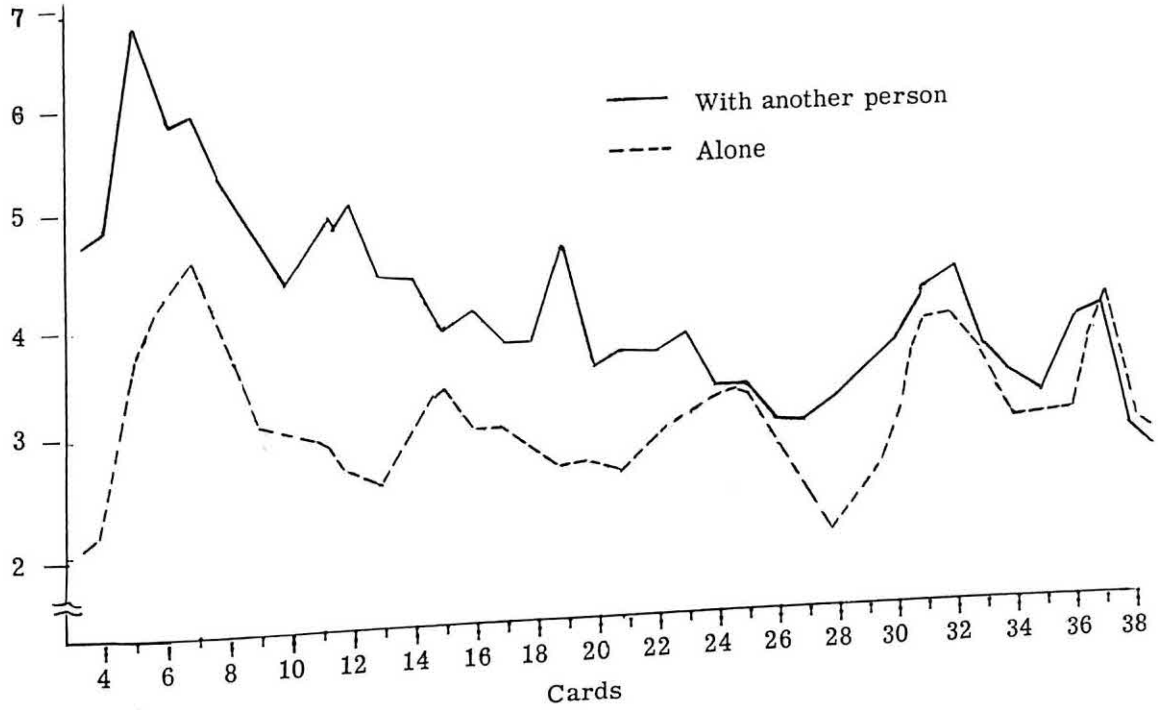


FIGURE 6

GSR AND ANXIETY LEVEL

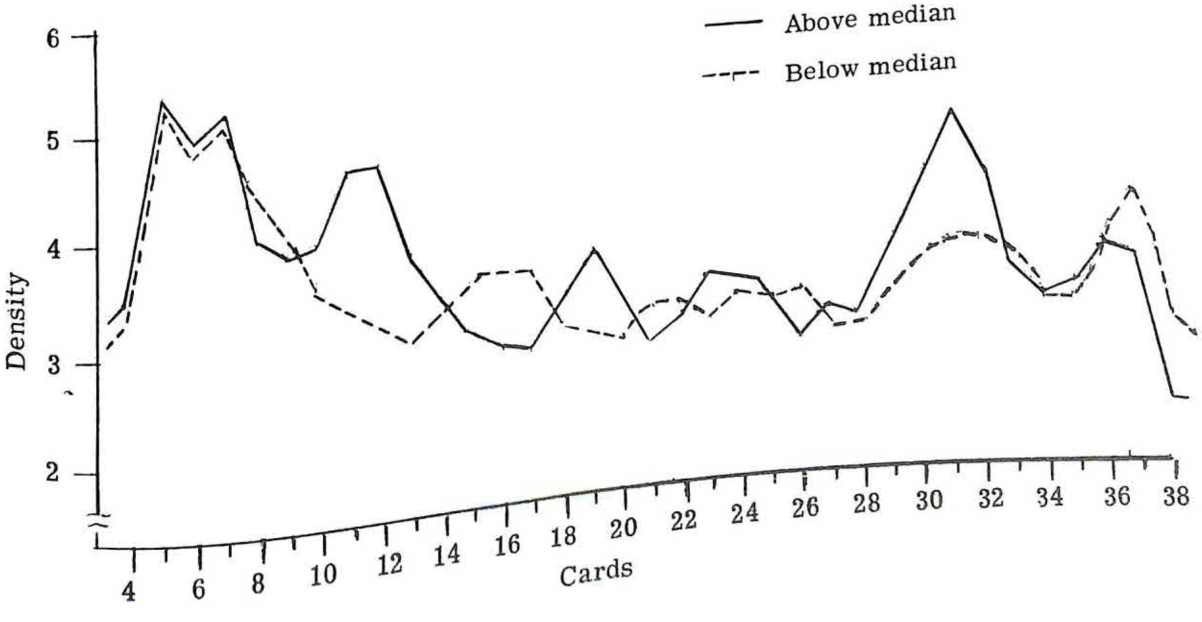
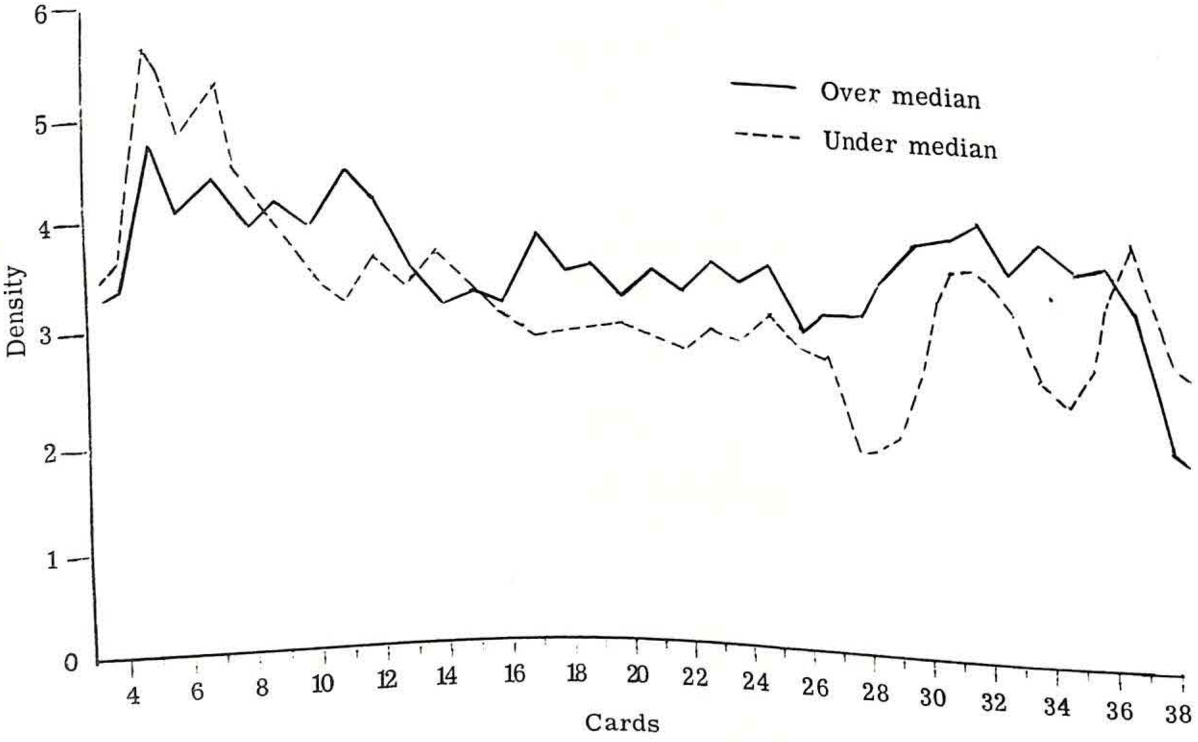


FIGURE 7

GSR AND ACADEMIC GRADES



to the frequency and amplitude variations within the specified limits and thereby retrieved information characterizing rate of division of change of galvanic skin response at any given time. By this means baseline drift was eliminated but only at the expense of measuring dermal resistance on an absolute scale.

Results

Figure 3 displays the mean galvanic skin response scores on the level of Interest Scale of the California Occupational Interest Inventory. The GSR reaction intensity is represented on the y axis and the time sequence on the x axis. The numbers on the x axis denote stimulus cards. The first three cards were used as "warm-up" and therefore are not included in the analysis of data.

Figure 3 illustrates a reasonably consistent higher intensity in GSR reactions on the part of the higher interest level group. This is of especial interest if one recalls that the Occupational Interest Inventory level of interest scale is fundamentally a measure of theoretical interests.

Figure 4 illustrates the mean galvanic skin response scores for subjects above and below the median on the Wonderlic Personnel Test (a gross index of mental ability).

The data in Figure 4 are consistent with that displayed in Figure 3. This would seem to indicate that the level of interest scale is concerned with theoretical interests only. From a combination of Figures 3 and 4, there seems to be a suggestion that the higher intelligence and theoretical interests are associated with basic physiological activity as reflected in GSR reactions.

Figure 5 illustrates the galvanic skin responses of subjects during the process of solving problems with and without the presence of another person. The experimental set-up called for the second person in the two person sessions to act as a "decoy" rather than serve as a subject. That is, the second person was "planted" to agree with the subject on certain problems and disagree on others. The routine was carried out in a nonsystematic manner.

The graph in Figure 5 illustrates basically what one would expect under the given set of circumstances. That is, the subjects showed a considerable original reaction to the presence and behavior of the "decoy" but apparently became acclimated to the situation. It is apparent from Figure 5 that the GSRs tended to get nearer and nearer together toward the end of the run.

Figure 6 displays the GSR graphs of subjects group above and below the median on the Taylor Anxiety Scale. It is apparent from Figure 6 that very little difference exists between the high and low anxiety. This somewhat surprising finding may be due to the fact that the total distribution for subjects on the Taylor Anxiety Scale clustered rather heavily about the median. If one used only extreme scores, different results might occur. In the present study the median Taylor score was equal to thirteen and the range extended from one to thirty-two.

Figure 7 illustrates the mean galvanic skin response scores for subjects grouped above and below the median on the basis of their overall academic grade average.

Figure 7 seems to be of inspectional interest only. The disparity between above and below median groups does not appear to be significant. The relatively consistent superiority of the above median group is suggestive, however, and possibly indicates some kind of trend. There was very little difference between the grades of the two groups, with a heavy cluster about the median. Therefore, the relatively small but consistent superiority of the above median group, coupled with the results of above median groups with reference to interests and intelligence, suggests that further study in this area might be profitable.

Summary

This study explored the possible relationship between galvanic skin responses and interests, intelligence, grouping, anxiety and grades. Definite trends seem to be indicated in regard to interests, intelligence, and grouping. No real trends are indicated in regard to anxiety and grades.

BIBLIOGRAPHY

1. Burch, Neil R. and Greiner, Theodore, "Drugs and Human Fatigue: GSR Parameters," *Journal of Psychology*, XLV (1958), pp. 3-10.
2. Cofer, Charles N., and others, "On the Significance of the Psychogalvanic Response as an Indicator of Reaction to Personality Test Items," *Journal of Psychology*, XXVII (1956), pp. 347-54.
3. Darrow, C. W. "The Galvanic Skin Reflex (Sweating) and Blood Pressure as Preparatory and Facilitative Functions," *Psychological Bulletin*, XXXIII (1936), p. 73.
4. Duffy, E., and Lacey, O. L., "Adaptation in Energy Mobilization: Changes in General Level of Palmar Skin Conductance," *Journal of Experimental Psychology*, XXXVI (1946), pp. 437-52.
5. Herr, V. V., and Kobler, F. J., "A Psychogalvanometric Test for Neuroticism," *Journal of Abnormal and Social Psychology*, XLVIII (1953), p. 410.
6. Herr, V. V., and Kobler, F. J., "Further Study of Psychogalvanometric Test for Neuroticism," *Journal of Clinical Psychology*,

XIII (1957), pp. 387-390.

7. Hugh, Halycon, and others, "The Relationship Between Certain Measurable Functions of Autonomic Nervous System Activity and Color Responses on the Rorschach Test," Journal of Clinical Psychology, VII (1951), pp. 244-49.
8. Jost, Hudson, The Use of Polygraphic Techniques in Psychophysiological Research and Clinical Psychology (Chicago: Associated Research, Inc., 1953).
9. Jost, Hudson, and Epstein, J. L. "The Rorschach as a Physiological Stress," Journal of Clinical Psychology, XII (1956), pp. 259-263.
10. Lacey, J. I., and others, "Autonomic Response Specificity," Psychosomatic Medicine, XV (1953), pp. 8-21.
11. Landis, C., and Hunt, W. A. "The Conscious Correlates of the GSR," Journal of Experimental Psychology, XVIII (1953), pp. 505-29.
12. Levy, E. Z., and others, "New Technique for Recording Skin Resistance Changes," Science, CXXVIII (1958), pp. 33-34.
13. Levy, E. Z., Applications of a New Technique for Recording Skin Resistances Changes, presented at the November 17, 1958 meeting of the American Rocket Society, New York.
14. Levy, E. Z., and others, The Use of Skin Resistance to Monitor States of Consciousness, presented to the 30th Annual Meeting of the Aero Medical Association, April 28, 1958, Los Angeles, California.
15. Levy, Jeanne R. "Changes in the Galvanic Skin Response Accompanying the Rorschach Test," Journal of Consulting Psychology, XIV (1950), pp. 128-33.
16. Lee, Clarence D. The Instrumental Detection of Deception (Springfield, Ill.: Charles H. Thomas, 1958).
17. McCleary, Robert A. "The Nature of the GSR," Psychological Bulletin, XLVII (1950), pp. 97-117.
18. McCleary, Robert A., and Lazarus, Richard H. "Autonomic Discrimination Without Awareness: an Interim Report," Journal of Personality, XVIII (1949), pp. 171-79.
19. McGinnies, Elliot, "Emotionality and Perceptual Defense," Psychological Review, LXVI (1949), pp. 244-51.
20. Martin, Barclay, "Galvanic Skin Conductance as a Function of Successive Interviews," Journal of Clinical Psychology, XII (1956), pp. 91-94.
21. Paintal, A. S. "Comparison of the GSRs of Normals and Psychotics," Journal of Experimental Psychology, IV (1951), pp. 425-428.
22. Ruff, G. E., and others, Biophysical and Physiological Considerations in the Interpretation of Changes in Skin Resistance, presented to the Aero Medical Panel Advisory Group for Aeronautical Research and Development, Paris, June 1958.
23. Schlosberg, Harold, "Three Dimensions of Emotion," Psychological Review, LXI (1954), pp. 61-81.
24. Schwartz, H. G., "Effect of Experimental Lesions of the Cortex on the Psychogalvanic Reflex in the Cat," Archives of Neurology and Psychiatry, XXXVIII (1937), pp. 308-320.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 2, December 1960)

ON THE USE OF Q-TECHNIQUE IN EDUCATIONAL EVALUATION AND RESEARCH

M. STEPHEN SHELDON
System Development Corporation, Santa Monica
A. GARTH SORENSON
University of California, Los Angeles

IF SCHOOLING is to be improved in a systematic fashion, it is first necessary to devise a means for appraising the learnings of individual students. This appraisal can of course be accomplished in part at the present time. The development of modern standardized achievement tests has made it possible to measure the progress of a student toward those educational goals which are defined mainly in terms of the acquisition of knowledge and skill. There are other objectives, however, which cannot be measured by standard achievement tests, but which are still considered vital. For example, it is commonly hoped that schooling will help a youngster to acquire a realistic awareness of his own capabilities and limitations, but it is usually difficult to determine the extent to which this awareness has been achieved. The lack of a generally accepted definition of "realistic awareness" does not make the objective seem any less important to teachers and the citizenry at large, nor does it seem likely that it will be abandoned. The problem facing educators, then, is to find a means of defining and observing the behavior signified by this and other similarly nebulous terms.

A procedure which has proven useful in areas of applied psychology is Stevenson's (1953) Q-technique. It is a special method of presenting test items and analyzing the results, and has been used a good deal in the appraisal of personality and in the measurement of attitudes and beliefs. Q-technique would appear to offer educators a means of dealing more systematically with some of their evaluation problems, and it is the purpose of this paper to present some illustrations of how the procedure can be used with groups of students. Four examples will be given. There will be a brief discussion of some of the problems which arise when Q-technique is used in the classroom. Some comments will be made about how the concepts of reliability and validity might be applied to Q-technique.

Example One: The Measurement of Personal Adjustment

The first study was undertaken to determine whether or not the twelfth grade students enrolled in a particular senior problems course increased in personal adjustment to a greater degree than a group of similar students who were not enrolled in the course. The index of adjustment used in this study was based on a correlation coefficient which was completed by comparing a student's Q sort description of his "ideal self" the person he would like to be—and his "perceived self"—his real or everyday self, again as described by the Q sort. Following Rogers (1951) and others, a high correlation between the perceived self and the ideal was taken as a sign of self-acceptance and personal adjustment. Conversely, a student who saw himself as being quite different from his ideal—the person he would like to be— was regarded as less well adjusted.

The first step in the study consisted of collecting an appropriate group of descriptive statements. To this end the students in several senior problems classes were asked to complete a number of unfinished sentences such as "My personality strengths are. . . "; "The ideal job should...'; "Good parents should. . . " Sixty self-descriptive statements were then written by the investigator, in consultation with the instructor of the senior problems course. Each statement related to an aspect of personal adjustment as the teacher employed that concept in the course, but was written in the idiom of the students, as revealed by their completions of the open-ended sentences. Care was taken to avoid ambiguity and to develop statements which were independent of one another. Each of the sixty statements was printed on a small numbered card, to permit easy sorting by the students. A set of cards was prepared for each student in the sample.

Three classes in senior problems, 67 pupils all taught by the same teacher, constituted the experimental group. A fourth class of 33 first semester seniors became the control group.

During the first week of the spring semester, each student sompleted "Self-Sort I". Each was given a set of statement cards, a record sheet on which to indicate the "score" assigned to a given statement, and a typed copy of directions. They were instructed to sort the statements into nine ranks in the way that best described themselves, and to assign a score to each statement according to its rank. The highest scores were assigned to the most descriptive; i. e., the highest ranking statements and the lowest score to those least descriptive. The range of scores (ranks) was 0 through 8. By specifying the number of statements to be given a particular scores—that is, assigned to that particular rank— the statements were forced into a quasi-normal distribution as shown in Table I. The students then recorded the score for each statement on the record sheet.

Two to four days later, the students completed "Ideal-Sort I", describing the person they would most like to be. At the end of the semester, they repeated these two steps, completing "Self-Sort II", and "Ideal-Sort II".

Two product-moment coefficients of correlation were computed for each student, one between Self-Sort I and Ideal-Sort I, and the second between Self-Sort II and Ideal Sort II. (It should be kept in mind that what was correlated was the rank or score assigned each descriptive statement on two different occasions.) This was not a laborious step, since the only computation necessary involved squaring the difference in rank (or score) assigned to each statement and summing the squares. The correlation was then read from a prepared chart. The record sheet was arranged to permit rapid comparison of scores and, by use of a desk calculator, any given coefficient of correlation could be computed in less than a minute.

The correlation coefficients were converted to z coefficients and the correlation between Self-Sort I and Ideal-Sort I was subtracted from the correlation between Self-Sort II and Ideal-Sort II. The difference was taken as the index of change in adjustment. The higher a positive difference, the greater the presumed change in self-acceptance. A negative difference presumably indicated that a student was less self-accepting, less well adjusted, than at the beginning of the course.

The mean differences of the individual experimental classes, and of the total experimental group were compared with the mean difference of the control group. The significance of the difference was computed, and the results are presented in Table II, which shows how and in what direction each group changed during the experimental period. It will be noted that, at the beginning, the control classes appeared to be better adjusted, but that at the end, they had lost ground in this respect.

On the other hand, the experimental group was lower at the beginning, but made some slight gains. These findings lead to the question of whether in our culture a decline in self-acceptance does in fact occur during the high school years. Does the imminent necessity for making important educational and vocational decisions produce a lack of security and self-assurance? Or are students acquiring ambitions and concepts of the ideal person which make them less satisfied with themselves? Q-technique could be used to investigate these and other important related questions.

Table III shows the frequency distribution of the coefficients of correlation between Self-Sort I and Self-Sort II for the experimental group. This correlation coefficient is a measure of the reliability of the students' self-concept over a period of five months. Considering the fact that the subjects were at an age when some change in self-perception is commonly believed to occur, plus the fact that they were in a class where some change in self-perception was being sought, the mean reliability of .65 would indicate that the self-concept of an adolescent is remarkably constant, even when measured by a relatively limited instrument.

Table IV compares the experimental and control classes with respect to change in personal adjustment, or self-acceptance. Ratios were calculated between the control group and each of the experimental classes, and between the control group and the experimental group as a whole. It will be noted that individually the experimental groups did not differ significantly from the control in their change in adjustment, but that the three experimental groups together did differ significantly at the .05 level of confidence. These results are inconclusive, but suggest that the senior problems class may not have been achieving its goal in the limited area investigated.

Example Two: Measuring Attitudes and Beliefs

The second study was conducted in a university course in curriculum development. This study investigated differences in beliefs about the public school curriculum between graduate students who were also school administrators and graduate students who were teachers. It was hypothesized that, when asked to sort statements about the public school curriculum, administrators would show a preference for a "subject-centered" curriculum, and teachers would show a preference for a "student-centered" curriculum.

As a first step, the instructor asked each student to list the four most important activities engaged in by pupils in the secondary school where he was employed. The resulting list of 220 activities was examined, duplications eliminated, and ambiguities clarified. The shortened list made

TABLE I

FREQUENCY DISTRIBUTION FOR THE SENIOR PROBLEMS SORT

										Least Like
Score	8	7	6	5	4	3	2	1	0	
Frequency	1	4	7	11	14	11	7	4	1	(n = 60)

TABLE II

MEANS AND STANDARD ERRORS OF CORRELATION COEFFICIENTS BETWEEN SELF-SORT I AND IDEAL-SORT I, AND SELF-SORT II AND IDEAL-SORT II FOR TWO GROUPS OF HIGH SCHOOL STUDENTS

	N	Correlation I Mean	Correlation I Sm	Correlation II Mean	Correlation II Sm
Control Group	37	.612	.048	.562	.065
Experimental I	21	.570	.074	.591	.077
Experimental II	22	.520	.080	.610	.071
Experimental III	24	.526	.058	.614	.076
Total Experimental	67	.538	.038	.616	.042

(The correlations have been converted to Fisher z coefficients).

TABLE III

DISTRIBUTION OF CORRELATIONS BETWEEN SELF-SORT I AND SELF-SORT II

Correlation	Frequency
.10 - .19	1
.20 - .29	2
.30 - .39	3
.40 - .49	5
.50 - .59	15
.60 - .69	15
.70 - .79	18
.8 - .89	8

(n = 67) (mean = .65)

TABLE IV

RATIOS BETWEEN THE MEANS ON THE MEASURE OF ADJUSTMENT FOR THE CONTROL AND THE EXPERIMENTAL GROUPS

Control with:	Before Measure		After Measure		Differences	
	t Ratio	Sig.	t Ratio	Sig.	t Ratio	Sig.
Experimental I	.580	*	.287	*	1.00	*
Experimental II	.988	*	.500	*	1.882	*
Experimental III	1.138	*	.521	*	1.858	*
Total Experimental	1.213	*	.566	*	1.984	.05

* Not statistically significant.

up a "parent" population of items. Drawing upon this list, the instructor then constructed thirty statements representative of a subject-centered curriculum, and thirty statements representative of a student-centered curriculum. Exemplifying subject-centered activities were: 1. studying the U. S. constitution, 2. studying geometry, 3. studying Latin. Among the student-centered activities were: 1. vocational guidance, 2. teaching communication skills, 3. learning to keep healthy.

Each of the graduate students was then instructed to sort the cards into a forced distribution as described in the first example. He assigned high scores to the activities he considered most important and low scores to those he considered least important. After sorting, he recorded a score (rank) for each item on the record form. This record form was so arranged that the instructor could independently total the scores assigned to student-centered items, and those assigned to subject-centered items. Thus he could quickly determine whether a student should be categorized as basically subject-centered or student-centered.

Since the N and the distribution were known and were the same for all sortings, the magnitude of difference to produce a critical ratio indicative to statistical significance could be calculated before administration of the items.

Intercorrelations of the sortings were computed for the five administrators and the fifteen teachers in the class. The resulting matrix showed that the administrators did not correlate any higher with each other than with the teachers. Thus the major hypothesis of this study was not supported.

Example Three: An Investigation in Educational Philosophy

Although this study was not completed, it illustrates how Q-techniques might be used to provide data regarding changes in a student's philosophy of education during a course, and data suggestive of the instructor's influence.

For purposes of this study, it was necessary that the investigator prepare an equal number of statements representing each of three philosophical positions: realism, idealism, and pragmatism. Each statement had to be mutually exclusive, i. e., had to characterize one ot the three positions but not the other two. A second dimension was built onto this group of items, cognitive vs. affective. That is to say, statements which dealt with metaphysics, epistemology, or logic were categorized as cognitive. Statements which dealt with axiology, ethics, or esthetics were classed as affective items. Adding this second dimension meant that six different categories of items (and an equal number of each) had to be prepared: cognitive-realistic, affective-realistic, cognitive-idealistic, affective-idealistic, cognitive-pragmatic, affective-pragmatic.

Because the boundaries between these philosophical positions are not clearly defined and generally agreed upon, the following steps were taken to secure items which would possess content validity. First, the investigator selected from the writings of philosophers and from text books in the philosophy of education, approximately 100 statements which appeared to represent one of the three philosophical positions but not the other two. These statements were listed in random order, copies of the list were prepared and given to five professors of philosophy, who had agreed to participate in the project. Each professor went through the list and independently marked each item to indicate which of the philosophical positions he believed it represented. A "cannot say" category was permitted where the professor found an item difficult to classify on any dimension. After computing the results of five such validations, the twelve most appropriate items were selected to represent each of the six categories. No item was used if it had been classified by any one of the professors as representative of more than one category on any dimension. Some of the items used, however, were agreed on by three or four professors, but placed in the "cannot say" category by the others. Thus a sample of 72 discriminating items was selected.

If this validated sample of propositions from philosophy were to be sorted by each student at the beginning of a course, say in the philosophy of education, and again at the end of the semester, number of important questions about the progress of the student in the course could be answered by following the procedures suggested below.

1. An analysis of variance of each student's scores would reveal (a) whether or not he held a consistent philosophical position (at least with respect to the statements provided for him to sort), (b) which of the three positions he favored, (c) whether his preferences were essentially for the cognitive, the affective, or both.
2. A correlation of the initial and final sorting would provide an index to the extent of change in the student's philosophy. The higher the correlation, the less the change.
3. Either or both of the student's sortings could be correlated with a sorting made by the instructor. If, for example, the final student-instructor correlation were significantly greater than the initial one, it would be evidence that the student's educational philosophy had moved closer to that of the instructor.

Example Four: Q-Techniques as an Instructional Tool

In this study, role-playing interviews were used as an instructional device in a counseling seminar. Q-technique was used to evaluate how well the counselor role was played. Forty descriptive statements were collected that could be used to describe both good and bad aspects of a counseling interview. Many of the statements were taken from Fiedler (1953) and reworded for application to a counseling situation. At the beginning of the class period, before the role-playing began, all students and the two attending faculty members were asked to sort the statements into the distribution indicated in Table V, to describe an ideal interview. The statements most characteristic of an ideal interview were assigned the highest scores, and those least characteristic were assigned the lowest scores. The score for each item was then recorded on a record form.

After a role-playing interview had been acted out, and before any discussion of it, the observers and participants were asked to sort the cards again, this time to describe the interview they had just observed. Any observers appraisal of the mock interview then could be obtained by correlating his Q sort description of the mock interview with his Q sort description of the ideal interview. The higher the correlation coefficient, the better his rating of the mock interview.

As an instructional device, the use of Q sorts presents many possibilities. It is especially useful for making students aware of problems of intra-observer and inter-observer reliability, of the influence of halo effect, and of the importance of operational definitions for such terms as "a good interview". For example, the use of Q-technique for the systematic appraisal of the mock interview produced a result very different from the global impressions of the observers. In the In the case of one student, the correlation between his Q-sort description of a mock interview which he had rated "good" on a global basis, and his Q-sort description of the "ideal" interview was only .04. This is a dramatic illustration of the questionable validity of the global impression of an untrained observer and provides one more bit of evidence for the case against global observations in general.

Another finding increased the student's awareness of the significance of operational definitions. It was discovered that the two instructors in the seminar, both professional counselors, one describing himself as "Rogerian" in philosophy, the other as "eclectic", correlated 76 in describing the ideal interview. (This of course is reminiscent of Fiedler's findings).

When all their sortings were intercorrelated, it was found that the students were in much closer agreement in describing the ideal interview, an ideational construct, than in describing an actual role-playing interview. It is suggested that, as a result of having had essentially the same courses and textbooks, they were responding to a stereotype in describing the ideal. The relatively low correlations between their descriptions of the real interview suggest that their verbalizations relating to counseling were still in the nature of rote learning and not yet understood in terms of the observable behavior of people. On the other hand, the low correlations suggest differences in perception, perhaps only some of which could be reduced by training— a difference which perhaps must be accounted for in terms of deeper psychological mechanisms relating to motivation.

Discussion and Evaluation

1. Is it feasible to employ Q-technique with groups of people? It is clear that concepts and procedures are being developed by clinical and counseling psychologists which, properly applied, can be extremely useful to educators in dealing with individual students. Since modern educators, however, must deal with large numbers of students and, usually, limitations in staff and budget, the new techniques will be more useful if they can be adapted for use with groups. The findings described above indicate that Q-technique can be readily and profitably applied to groups. In these four studies, once the descriptive statements had been collected (and of course that is the most crucial step in the entire procedure), sets of items were quickly prepared, using standard duplicating processes and a paper cutter, and were found to be satisfactory in both small and large classes. It was found that, at the high school and college level, the mimeographed instructions for sorting and the recording of scores were adequate. At the college level, for example, with one page of written directions, only one of 122 students needed additional instruction. At the high school level, when directions were read to the students, very few of them made errors in their first sortings, and none in subsequent sortings.

2. Using the same items, would not a Likert-type scale be more useful? When the responses of two or more people are to be compared, or the responses of one person made on two or more occasions, Q-technique offers at least two advantages. The most obvious is the comparative ease of statistical treatment. To correlate the responses of two people to a 60-item Likert-type scale requires about twenty minutes, using a modern desk calculator. In contrast, two 60-item Q-sortings have been recorded on appropriate record forms, and with the use of prepared tables, can be correlated in 45 seconds using only an adding machine. A second advantage over the Likert-type scale is the larger number of items that can be presented. Each forced choice is considered equivalent to one item.

3. Is Q sort subject to faking? The degree to which an interest or personality inventory can be faked depends in large part on the obviousness of

TABLE V

FREQUENCY DISTRIBUTION FOR THE COUNSELING SORT

	Most Desirable						Least Desirable	
Score	6	5	4	3	2	1	0	
Frequency	2	5	8	10	8	5	2	(n = 40)

the items, on differences in social acceptability among the items, and on whether or not a forced-choice technique is used. Unless a person who writes the items possesses unusual skill, it seems doubtful that his Q items will be as subtle as empirically derived scales developed by professional test makers.

With respect to social acceptability, however, it would appear that Q-technique may offer distinct advantages. In many personality inventories, the "adjustment" score depends on the number of unfavorable characteristics a subject is willing to admit he possesses. But psychological adjustment, when defined by Rogers and others in terms of self-acceptance, depends on congruence between the perceived self and the ideal self. If this concept is sound, Q-technique provides a powerful tool indeed. It would take a discriminating and psychologically sophisticated person to perceive that the index of adjustment does not depend on the self-description, but rather on self-acceptance. Although there are problems of social acceptability inherent in Q-technique, these can be dealt with by those methods developed recently by Edwards (1957). Since Q sorting presents the subject with a more complex task than does the usual forced-choice test, it would presumably be harder to fake.

4. At what educational level is Q-technique appropriate? As reported above, it has been used successfully with high school students. If sufficient care is taken to write clear directions, and to use words of the appropriate level of difficulty, it seems reasonable to believe that Q sorts can be developed for use at the 7th grade level.

5. Where do the descriptive statements come from? Obviously the first step in writing items is to give careful thought to the "dimensions" to be represented in the Q sort. These will be determined by the questions the investigator wants to answer. Once the writer has decided on the "structure" of his instrument, then, by asking "open-ended" questions of a sample of individuals from the group he intends to investigate, he can acquire a pool of items in the idiom of the subjects themselves.

If the behavior, or concepts, or beliefs under scrutiny require that the items possess content validity as in the case of the "philosophy sort" described above, then the writings of the acknowledged experts in a given field may be used as a source of items.

6. Can Q-technique be used to measure change or achievement? By correlating the sortings made by one person on two different occasions, an investigator can find out whether or not change has occurred in a subject. If, for example, a science teacher wished to measure the changes in attitude toward science which occurred in his class, four steps would be involved: 1) The instrument (items and directions) would be constructed. 2) The instructor would sort the items to describe the ideal attitude, the one he would want his students to acquire, and would record his sorting on a record form. 3) the students would sort the items at the beginning of the course to describe his attitude toward science. 4) The student would sort the items at the end of the course to describe his attitude. By correlating the student's first sort with the ideal, the instructor could determine degree of congruence. By correlating the second sort with the ideal, an instructor could get a second measure of attitude. By comparing the two correlations, he could determine whether change had occurred and in what direction.

7. Are the statistical procedures laborious? Two statistical shortcuts have been described which can be employed by a statistical clerk, or by someone with a minimum of statistical knowledge. If any dimension of a sample of items is dichotomous, the necessary t-ratios for differentiating the categories on a statistically significant basis can be computed before administration. Such t-ratios may be calculated for any desired level of confidence and applied to all sortings of the same items. This permits classification of the subject in terms of the variable as soon as the scores assigned to the items are summed.

When many correlations are to be computed between sortings, a conversion table or graph (Cohen, 1957) may be constructed. This permits the reading of a correlation coefficient directly from the summation of the squared differences between the scores (ranks) assigned each item of the sort.

Summary

Q-technique permits appraisal of the beliefs, attitudes, values and self-concepts of individual students in a way not provided for by conventional psychological tests and inventories. Consequently it offers educators a means of solving problems of measurement which could not heretofore be dealt with. This paper has provided some illustrations and suggestions regarding the use of Q-techniques with groups of students. The studies reported were mainly exploratory in nature and by no means exhaust the applications to be made. Nor do they reflect the more profound possibilities of Q methodology as envisioned by its major author, Stephenson (1953). The experience with these studies, however, resulted in two major conclusions: Q-technique can be applied in many intriguing ways to important problems of educational measurement. It is highly feasible to use Q-technique with groups of students. Indeed, for some purposes, the technique is more flexible and less time-consuming than conventional methods.

REFERENCES

1. Cohen, J. "An Aid in the Computation of Correlations Based on Q-Sorts," *Psychological Bulletin*, LIV (1957), 138-139.
2. Edwards, A. *Social Desirability Variable in Personality Assessment*, New York: Appleton-Century-Croft, 1957.
3. Fiedler, F. E. "Quantitative Studies on the Role of the Therapist's Feelings toward his Patients," in O. H. Mowrer, ed., *Psychotherapy: Theory and Research*, New York: Ronald Press Company, 1953
4. Rogers, C. R. *Client Centered Therapy*, Cambridge, Mass.: Riverside Press, 1951.
5. Stephenson, William. *The Study of Behavior*, Chicago: University of Chicago Press, 1953.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 2, December 1960)

THE USEFULNESS OF A TWO-WAY ANALYSIS OF WISC SUB-TESTS IN THE DIAGNOSIS OF REMEDIAL READING PROBLEMS

LYNNE SCHELLBERG HIRST
University of Wisconsin

THE BASIC purpose of this paper is to explore the usefulness of a two-way method of analyzing the sub-test scaled scores of the Wechsler Intelligence Scale for Children (WISC) in the diagnosis of remedial reading problems, so as to facilitate a more accurate appraisal of intellectual functioning in regard to retardation in reading. The second purpose was to note any distinctive test patterns in the WISC sub-test scaled scores that are characteristic of severely and mildly retarded readers.

The type of analysis employed in this paper is the "scatter analysis", which refers essentially to an examination of the sub-test scaled scores or raw scores in relation to the particular mean score that is being used.

One of the most common methods of analyzing the scatter of sub-test scaled scores on the WISC is to relate these scores to the scaled score of 10, which is the scaled value assigned to the mean raw score for each sub-test in the standardization of the test for each age group. Three scaled scores represent a sigma in the distribution of raw scores for each age group's sub-tests. However, for the purpose of this paper, two scaled scores, or the equivalent of two-thirds of a sigma, will be referred to as being significant. To support this, this study is influenced by the usual clinical practice of accepting this interpretation.

Relating the scaled scores to the scaled score mean of 10 gives the individual's weaknesses and strengths in relation to the test's equivalent age group. This method ignores, however, the related weaknesses and strengths within the individual per se. For example, the majority of the scaled scores for bright children fall in the proximity of the scaled score of 12, and few of their relative weaknesses fall below 8, so that their relative weaknesses and strengths are difficult to ascertain. This is also true in reference to dull children since a high percentage of their scaled scores fall close to the scaled score of 8 and few of their relative strengths fall above the scaled score of 12. Thus, a complete picture of the individual's strengths and weaknesses is not presented by this method alone.

A second method of analyzing the scatter of the sub-test scaled scores is to relate each of these scores to the individual's own scaled score mean. However, the difficulty with this method is that if a person with a mean of 5 scored a 7 on a sub-test, this scaled score of 7 would be considered significantly high. Thus the relationship of the individual's weaknesses and strengths to the scaled score mean of the test's equivalent standardized age group is overlooked.

While the use of either method of scaled score would appear to have particular advantages and disadvantages, if the two methods of analysis could be related to each other, a valuable means of analysis of scatter might be provided.

Miss Ella Bullis of the University of Wisconsin Reading Clinic has developed a way of analyzing the scaled scores that account for both methods. On one dimension of her two-way analysis scale are the scaled score mean of 10 and the deviations from this mean so that the relationship of the individual's scores to his equivalent age group's scores is taken into account. On the other dimension are the individual's own mean and the deviation scores from it so that the subject's own weaknesses and strengths are considered. This two-way analysis is used in this paper.

Methodology and Population of the Study

The two-way analysis employed in this study can best be explained by Figure 1 (see top of next page).

Across the top from left to right are grouped three categories of scaled sub-test scores. It is assumed that two scaled scores, representing two-thirds of a sigma from the mean of 10, is a significant deviation from the mean. Thus the scaled scores of 9, 10, and 11 are classified in the chart as "average"; the scaled scores of 12 or above as "high" and the scaled scores of 8 or below as "low".

The other axis consists of the individual's mean scaled scores. Two scaled scores are likewise assumed to represent a significant deviation in the distribution of scores from the individual's mean

FIGURE 1

THE TWO-WAY ANALYSIS

Significant Deviations from the Scaled Score Mean of 10

	High 12+	Average 11, 10, 0	Low 8-
Above			
Individual's Mean			
Below			

score.

This two-way method of analyzing sub-test scaled scores was employed on thirty remedial reading cases from the 1958 and 1959 Summer Reading Clinic cases at the University of Wisconsin. The subjects for this study were selected on the following criteria: 1) Their full scale IQ must be 89 or above; 2) their reading level as measured by the Chicago Silent Reading Test, Form 2, and/or the Gates Advanced Primary test must be six months or more below their mental age; 3) their chronological age must be between 8-0 and 13-6.

The mean Verbal Scale IQ for the thirty subjects was 100. 5, the mean Performance IQ was 111. 6, and the Full Scale IQ was 109. 26. The eleven point difference between the Verbal and Performance Scale IQ's was found not to be significant. The average CA for the group was 10-3, and the total range of reading disability for the group was found to be from seven months to four years and eleven months.

The group of thirty children was then divided according to the degree of reading disability. One group consisted of those individuals who had a reading disability of two or more years and were called the severe reading disability group. The second group, called the mild reading disability group, consisted of those individuals who had a reading disability of less than two years.

A sample of the two-way analysis method as applied to one of the reading cases is shown on the following page.

Results

The analysis of the results are divided into three sections: the total group's results, the mild reading disability group's (MR's) results, and the severe reading group's (SR's) results. The purpose of this section of the paper is, first of all, to examine the percentage of individual's own mean and the scaled score mean of 10. These means will be referred to from now on as the two-way analysis means, and deviations of two scaled scores from both means is considered significant. Second, it is the purpose to note any distinctive test patterns in the above groups' sub-test scaled scores that are characteristic of these groups.

Total Group's Results

Upon examining the results of the two-way analysis method as applied to the thirty remedial readers (Table I), one sees that 50% and 47% of the group's individuals scored significantly above the two-way analysis means on the Picture Completion and Picture Arrangement sub-tests, respectively.

In addition, approximately one-fifth of the group scored significantly above the two-way analysis means on the Block Design and the Object Assembly sub-tests.

In respect to those scores on which the group performed poorly, it was found that 38%, 37%, and 37% of the total group scored significantly below the two-way analysis means on the Digit Span, Arithmetic, and Coding sub-tests as did 23% of the group on the Vocabulary sub-test.

Mild Reading Group's Results

Fifty percent of the MR's, that is, those individuals whose disability in reading is less than two years (Table II), scored significantly above the two-way analysis means on the Picture Completion and Picture Arrangement sub-tests. Also

SAMPLE OF TWO-WAY ANALYSIS

Name - P. S.

	Scaled Score	I Q
CA 9-2		
Verbal Scale IQ	50	100
Performance Scale IQ	51	101
Full Scale IQ	101	101*

		Scaled Score Scatter	
	High	Average	Low
Above	Information (15) Picture Arrangement (14) Object Assembly (13)		
Individual's Mean	Comprehension (12)	Similarities (10) Coding (11)	
Below			Arithmetic (5) Vocabulary (8) Picture Completion (6) Block Design (7)

*Digit Span omitted

TABLE I

THE PERCENTAGE OF THE TOTAL GROUP'S SUB-TEST SCALED SCORES THAT DEVIATE SIGNIFICANTLY ABOVE AND BELOW THE TWO-WAY ANALYSIS SCORES

	High	Low
Information	10%	7%
Comprehension	17%	13%
Arithmetic	3%	37%
Similarities	13%	17%
Vocabulary	3%	23%
Digit Span	0	38%*
Picture Completion	50%	10%
Picture Arrangement	47%	13%
Block Design	23%	13%
Object Assembly	20%	3%
Coding	13%	37%

*For 21 of the subjects

TABLE II

THE PERCENTAGE OF THE MILD READING DISABILITY GROUP'S SUB-TEST SCALED SCORES THAT DEVIATED SIGNIFICANTLY ABOVE AND BELOW THE TWO-WAY ANALYSIS MEAN

	High	Low
Information	11%	6%
Comprehension	22%	17%
Arithmetic	5%	28%
Similarities	11%	11%
Vocabulary	5%	17%
Digit Span	0	38%*
Picture Completion	56%	11%
Picture Arrangement	50%	10%
Block Design	33%	22%
Object Assembly	5%	5%
Coding	11%	28%

*For 18 of the subjects

TABLE III

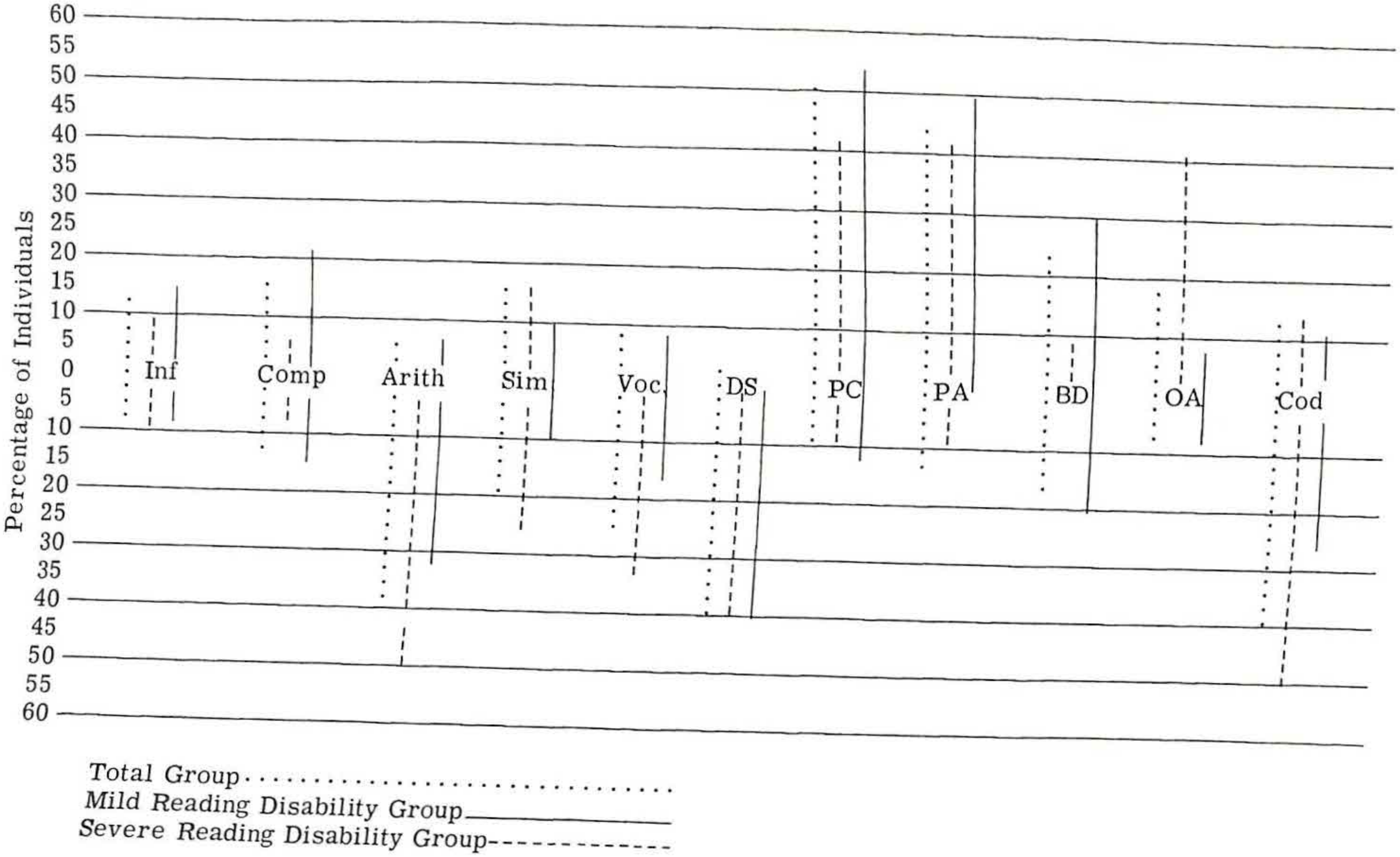
THE PERCENTAGE OF THE SEVERE READING DISABILITY GROUP'S SUB-TEST SCALED SCORES THAT DEVIATE SIGNIFICANTLY ABOVE AND BELOW THE TWO-WAY ANALYSIS MEAN

	High	Low
Information	8%	8%
comprehension	8%	8%
Arithmetic	0	50%
Similarities	12%	25%
Vocabulary	0	33%
Digit Span	0	38%*
Picture Completion	42%	8%
Picture Arrangement	42%	8%
Block Design	82%	0
Object Assembly	42%	0
Coding	17%	50%

*For 9 of the subjects

FIGURE 2

A GRAPHIC DESCRIPTION OF THE DIFFERENCES BETWEEN THE THREE READING DISABILITY GROUPS



22% of this same group scored significantly above the two-way analysis means on the Comprehension sub-test.

Twenty-eight of the MR's deviated significantly below the two-way analysis means on the Coding and Arithmetic.

On the Block Design sub-test, approximately the same percentage of individuals deviated significantly above as they did below the two-way analysis means. Noticeable, also, was that on the Object Assembly sub-test very few individuals deviated above or below the two-way analysis means.

Severe Reading Disability Group's Results

Forty-two percent of the SR's, that is, those individuals with a reading disability of two or more years, scored significantly above the two-way analysis means on the Picture Completion, Object Assembly, and Picture Arrangement sub-test, respectively. The percentage of individuals who scored significantly above the two-way analysis means is surprising in view of the fact that most of the mild reading disability cases scaled scores did not deviate from these means on this sub-test (Table III).

Fifty percent of this group scored significantly below the two-way analysis means on the Arithmetic and Coding sub-tests. Also 38%, 33%, and 25% of this group scored significantly below the two-way analysis means on Digit Span, Vocabulary, and Similarities sub-tests.

A very small percentage of the SR's scored above or below the two-way analysis means on the Information, Comprehension, and Block Design sub-tests.

Summary

Figure 2 gives a graphic description of the differences between the three reading disability groups.

In summary, then, the differences between the two groups are striking in regard to certain sub-tests. The Object Assembly sub-test is one, for example, on which only 5% of the MR's scored significantly above the two-way analysis means; however, 42% of the SR's scored significantly above these means. Also, the percentage of individuals of the SR group who scored below the two-way analysis means on the Arithmetic and Digit Symbol sub-tests was much higher than the percentage of the MR group. In addition, a smaller percentage of the SR's scored above the two-way analysis means on the Picture Completion and the Picture Arrangement sub-tests. A higher percentage of the SR's scored below the two-way analysis means on the Vocabulary and Similarities sub-tests. Also while approximately the same percentage of individuals in the MR group scored significantly above and below the two-way analysis means, most of the individuals in the SR group scores did not deviate from the two-way analysis means.

Thus, the characteristic pattern for the total reading disability group in general would be high Picture Completion and Picture Arrangement sub-test scaled scores and possibly high Block Design and Object Assembly sub-test scaled scores. Also low Digit Span, Arithmetic, and Coding sub-tests and possibly Vocabulary might be included.

For the MR's, High Picture Completion and Picture Arrangement sub-test scaled scores with possibly high or low Block Design in sub-test scaled scores would be expected along with low Coding, Arithmetic, and Digit Span sub-test scores.

For the SR's, high scaled scores would be expected on the Picture Completion, Object Assembly, and Picture Arrangement sub-tests. Also low Arithmetic, Coding, and Digit Span sub-test scaled scores along with possibly low Vocabulary and Similarities sub-test scaled scores could be anticipated.

Interpretation and Implications for Further Research

Now that we have detected certain sub-test patterns that are characteristic for mild and severe reading disability cases, what we do not know is the significance of these patterns. What we need is further research to test what these sub-tests really means. If the Picture Completion test means that the individual is able to detect small differences in picture forms and the Picture Arrangement test measures the ability to see pictures in sequence, the implications for the teaching devices and measurements now being used in the field of reading would have to be reconsidered. If the Coding sub-test measures the rate of new learning or the visual-motor ability of individuals, the implications of this test could be widely spread. However, before we can use these sub-tests in accurately appraising the intellectual functioning in regard to retardation in reading, we must know what these tests represent. It would also be valuable to explore the validity of this two-way method in the diagnosis of certain populations such as the mentally deficient and the emotionally disturbed. For example, since clinicians are now using parts of this analysis, it would be interesting to see the results of this method applied to different types of cases which they diagnose. In this way new relationships might be established between certain types of disorders and the scatter of the WISC sub-test scaled scores.

The two-dimensional approach used in the analysis of sub-test scatter in this paper appears to present a more precise and meaningful analysis of intellectual functioning than do uni-dimensional analyses of scatter in such studies as Atlus (1), Graham (2), Hurst and Portenier (3), and Wilke (6). Since it provides a dual reference for judging the extent of deviation on sub-test scores, greater confidence can be attributed to the nature and significance of the sub-test pattern.

REFERENCES

1. Atlus, Grace F. "A WISC Profile for Retarded Readers," Journal of Consulting Psychology, XVI (1952), pp. 155-56.
2. Graham, E. E. "Wechsler-Bellevue and WISC Scattergrams of Unsuccessful Readers," Journal of Consulting Psychology, XVI (1952), pp. 268-71.
3. Hurst, Wilma E., and Portenier, Lillian G. "The Wechsler Scale as a Diagnostic Tool for Cases of Reading Disability," Journal of Colorado-Wyoming Academy of Science, IV (1951), p. 71
4. Wechsler, D. The Measurement of Adult Intelligence (New York: Williams and Wilkins Co., 1944), pp. 153-58.
5. Wechsler, D. Wechsler Intelligence Scale for Children (New York: Psychological Corporation, 1949), pp. 7-16.
6. Wilke, Richard E. Study of Eleven Reading Clinic Cases and WISC Sub-Testing Analysis in Relationship to Reading Retardation, unpublished seminar report, University of Wisconsin, 1957.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 2, December 1960)

PRE-TEST AND RE-TEST SCORES IN RETENTION CALCULATION*

ELLIS BEECHER LITTLE
University of Illinois, Chicago Undergraduate Division

EXAMINATION OF several studies in retention will reveal certain methodological procedures characteristic of the "conventional" method for the calculation of percent retention of meaningful material.

The calculation of percent retention is accomplished in the "conventional" method by administering a pre-test to the students who are to take the course, in order to determine how much course information is known in advance of course presentation. The student then takes the course. At the end of the course a final examination is administered, and in order to determine how much the student learned in the course, his pre-test score is subtracted from his final examination score. At some later time--six months, one year, two years, five years--a re-test is given the student. The score on this re-test minus the original pre-test score equals a score representing the amount of information retained of information learned in the course. The percent retention is then calculated by dividing the score for material retained of material learned in the course by the score for material learned in the course.

Research evidence indicates, however, that forgetting is a function of activities prior to the learning, as well as activities subsequent to the learning. We may assume that this applies not only to learning in a course, but also to material of the course that was known before the course was taken. In other words, a score on a course pre-test would be expected to change because of activities prior to, and subsequent to the learning represented by such a score. The assumption may be made, therefore, that a score for items answered correctly on a pre-test will show change (drop) for those same items on the post-test and on the later or delayed re-test. This is an accepted truism concerning material learned in a course and should apply to material of the course that was known or learned before the course was started.

Research evidence indicates also that under certain conditions the passage of time results in an improvement, rather than a decrement, in performance. Although Ballard's study may be criticized for its inadequate control of practice and rehearsal, it does suggest that improvement may take place between course completion and administration of a later or delayed re-test. The assumption may be made, therefore, that between the time of post-test administration at the time of course completion, and the administration of the later or delayed re-test, some learning will take place. We need not go into the problem of whether the learning was delayed course learning, new learning entirely, or a combination of both.

Applying the above research evidence to the calculation of percent retention we have a pre-test score which is comprised of the items answered correctly at the time of pre-test administration; a post-test score which is comprised of two scores: 1) a score representing what is retained of what was known before the course was taken, 2) a score representing what is known as a result of course learning; and a delayed re-test score which is made up of three scores: 1) a score representing what is retained of what was known before the course was taken, 2) a score representing what is retained of what was learned in the course, 3) a score representing what is learned after the course is completed.

The Problem

Three problems present themselves at this point: 1) Will the pre-test score remain constant over a period of time? 2) Is learning taking place between the time of course completion and the time of re-test administration? 3) If the pre-test score does not remain constant and if there is learning taking place between the time of course completion and re-test administration, will these changes significantly affect the calculation of percent retention of meaningful material?

---

* This article is based on an Ed. D. thesis presented to the Graduate Faculty of the University of Illinois. The writer is grateful to Professor R. Will Burnett, advisor and chairman of his special committee, and to Professor J. Thomas Hastings for the invaluable assistance given him in preparing and writing the thesis.

Hypotheses

a. That an item count of the items answered correctly on the pre-test, when compared to a count of those same items on both the post-test and delayed re-test, will show a drop on both the post-test and delayed re-test from what it was on the pre-test.

b. That an item count of items found to be correct on the delayed re-test when compared item for item with items found to be correct on the pre-test and post-test will tend to show that learning has taken place between the time of course completion and the administration of the delayed re-test.

The above hypotheses were then used in an experimental determination of percent retention of meaningful material as calculated by the formula:

1. Score for correct pre-test
items still correct on re-test.
Re-test score-
2. Score for material learned
since course completion.
× 100 = percent retention

Post-test score - 1. Score for correct pre-test
items still correct on the
post-test.

A comparison was then made between the experimental percent retention as obtained from the above formula, and the "conventional" percent retention as calculated by the "conventional" formula:

$$\frac{\text{Re-test score - Pre-test score}}{\text{Post-test score - Pre-test score}} \times 100 = \text{Percent Retention}$$

Method and Procedure

Two groups of college students, twenty-six in one group and thirty-eight in the other, were tested by a machine-graded type examination for recognition of biology material learned before taking a general biology course (pre-test), for recognition of material learned during the time of the course (post-test), and for recognition of material retained of material learned during the time of the course (re-test). In an attempt to insure a normal population, the students' sten scores on the Michigan Vocabulary test were used as a basis for student selection.

For purposes of this study, test items dealing with facts in beginning biology were selected from a final examination. This final examination was compiled by a jury of several members of the Department of Biological Science, Chicago Undergraduate Division, University of Illinois. The items selected for the experimental examinations were selected on the basis of whether or not factual information was asked for, and whether or not they could be considered recognition-type items.

The reliabilities of the tests were calculated on the immediate retention or post-test and were calculated in terms of the Pearson product-moment method of correlation, using the split-half technique. The coefficients were then corrected by the Spearman-Brown formula. The reliability coefficients for the February and September test items are .88 and .82 respectively. The probable errors are not more than plus or minus .034 and plus or minus .015 respectively.

The claim may be made that the two tests used have validity since there was a high degree of concurrence of judges or experts in biology as to what the correct answers ought to be.

Scoring

According to Damrin, there are three ways of scoring a test: 1) the empirical method of scoring, 2) the logical method of scoring, and 3) the factorial method of scoring.

The test used for this study was scored by the logical method. We were interested in securing an evaluation of what the student ought to know or how he ought to respond about or to a group of selected items in beginning biology. The keys to these tests, therefore, represented a preconceived pattern of responses to select stimuli in beginning biology. Since the keys represented a preconceived pattern of responses, it was thought best to submit the test to three experts in the field of biology and have them each make a key. The three keys were then compared and only items upon which the experts agreed were used as test items in this study.

Results

As indicated above, hypothesis a) divides itself into two sections, 1) that there will be a significant drop in the post-test score for those items answered correctly on the pre-test, and 2) that there will be a significant drop in the re-test score for items answered correctly on the pretest.

As shown in Table I, there is an average drop of -2.8 items and -3.4 items from pre-test to post-test when scores are calculated only on items correct on the pre-test. The coefficient of correlation between the pre-test scores and the scores for the same correct items on the post-test is .95 for the February group and .86 for the September group respectively. The significance ratio, using the coefficients of correlation in the formula Eml - m2 is 4.6 for the February group

TABLE I

AVERAGE DIFFERENCE BETWEEN SCORES FOR ITEMS ANSWERED CORRECTLY ON THE PRE-TEST AND FOR THE SAME ITEMS ON THE POST-TEST AND RE-TEST

Group	Ave. Dif. Between Pre-test and Post-test	T	Ave. Dif. Between Pre-test and Re-test	T
February	-2.8	4.6	-4.34	4.5
September	-3.4	5.0	-4.47	5.1

TABLE II

SCORES FOR MATERIAL LEARNED SINCE COURSE COMPLETION

Group	Ave. No. of Items Answered Correctly on Re-test and not Answered on Pre or Post-test	T
February	+4.1	4.2
September	+2.8	4.1

which gives a probability of 1 in 473,000 that the difference could occur by chance. This is significant at a confidence level way beyond one percent For the September group, the significance ratio is 5.0 which gives a probability of 1 in 3,490,000 that the difference could occur by chance. This is significant at a confidence level approaching absolute certainty.

As is also shown in Table I, the average difference between items correctly answered on the pre-test that were also answered correctly on the re-test, are score drops of -4.34 and -4.47 items. The coefficients of correlation between the original pre-test scores and the number of those same items correct on the re-test for the February and September groups are .87 and .79 respectively. When these coefficients of correlation are used in calculating EM1-M2, the significance ratios are found to be 4.5 and 5.1 which give probabilities of 1 in 294,000 and 1 in 4,000,000 that the differences could occur by chance.

Hypothesis b) states that an itemcount of items found to be correct on the re-test examination when compared item for item with items found to be correct on the post-test, will tend to show that learning has taken place between the test of course completion and administration of the re-test examination.

As shown in Table II, the average differences between an item count of items found to be correct on the re-test examination and items found to be correct on the post-test examination are +4.1 items for the February group, and +2.8 items for the September group. The coefficients of correlation between the scores making up these averages are .85 and .88 respectively. Using these coefficients of correlation to calculate EM1 -M2, the significance ratios of 4.2 and 5.1 are obtained. These ratios give probabilities of 1 in 75,200, and 1 in 48,300 that the differences could occur by chance.

Since the foregoing tests for significance of differences have all been significant at the one percent level of confidence or beyond, we may assume that there are other than chance factors responsible for the differences in scores herein tested. We may, therefore, accept hypothesis b), and state that an item count of items found to be correct on the re-test examination, when compared item for item with items found to be correct on the post-test, will show that learning has taken place between the time of course completion and the time of the administration of the re-test.

As shown in Table III, when the mean percent retention is calculated using the corrected scores of both hypothesis a) and hypothesis b) and when this mean percent retention is compared to the mean percent retention as calculated using the scores as they would be used in the "conventional" method, it is found that the mean percent retention as calculated by the experimental method is less than the mean percent retention as calculated by the "conventional" method.

For the February group, the difference between the mean percent retention as calculated by the "conventional" method and the mean percent retention as calculated by the experimental method is 71.02 percent ("conventional") minus 64.44 percent (experimental) or -6.58 percent. For the September group, the difference between the mean percent retention as calculated by the "conventional" method and the mean percent retention as calculated by the experimental method introduced in this study is 74.5 percent ("conventional") minus 70.3 percent (experimental) or a mean difference of -4.2 percent. For T, the significance ratios are .96 and .62 respectively for the -6.58 percent and -4.2 percent differences and give probabilities of 1 in 4.9 and 1 in 2.7 that the differences could occur by chance.

Discussion

The Stability of the Pre-test Score. The results of part 1) of hypothesis a) proved significant. There are drops in score of -2.8 and -3.4 items from the average number of items correct on the pre-test to the average number of those same items correct on the post-test. Thus the experimental corrected pre-test score when subtracted from the post-test score results in an increase in the score for the amount of material learned in the course. This is to say that the score for the amount of material learned in the course is greater when the score for the pre-test is adjusted as described in this study than it is when the "conventional" unadjusted pre-test score is used.

The results of part 2) of hypothesis a) proved significant. There are drops in score of -4.34 and -4.47 items from the average number of items correct on the pre-test to the average number of those same items that were correct on the re-test. Thus, the experimental, corrected pre-test score subtracted from the re-test score results in an increase in the score for the amount of material retained of material learned in the course.

The results for hypothesis b) proved to be significant. The average score differences of 4.1 and 2.8 items, for material learned between course completion and the time of the re-test, when used as indicated in this study, decreases the scores for material retained of material learned in the course.

The above score changes bring about the following changes in percent retention:

1. When the score for material learned in the course is increased, the percent retention of material retained of material learned in the course is decreased.
2. When the score for the amount of material retained of the material learned in the course is

TABLE III

PERCENT RETENTION AS CALCULATED BY "CONVENTIONAL" AND EXPERIMENTAL METHODS, ALSO SHOWING MEAN DIFFERENCE BETWEEN METHODS AND SIGNIFICANCE RATIO FOR DIFFERENCES

Group	Mean Percent Retention by "Conventional" Calculation	Mean Percent Retention by Experimental Calculation	Mean Difference	T
February	71.02%	64.44%	-6.58%	.96
September	74.50%	70.30%	-4.20%	.62

increased, the percent retention of material retained of material learned in the course is increased.

3. When the score for the amount of material retained of the material learned in the course is increased (as it is when material learned since course completion is taken into consideration) the percent retention of material retained of material learned in the course is decreased.

4. These combined differences, some of them occurring in opposite directions, balance each other out to some extent. The February group's experimental percent retention of material learned in the course was -6.58% less on the average than its "percent retention" as calculated by the "conventional" method. The September group's experimental percent retention of material learned in the course was -4.10% less on the average than its percent retention as calculated by the "conventional" method. It is interesting to note that the average percent retention as calculated by the experimental method is less in each case for both the February and September groups.

However, the calculation of the average percent retention by the experimental method introduced in this study proved not to be statistically significant from the average percent retention as calculated by the "conventional" method.

Summary

The following two hypotheses were tested:

a. 1) That an item count of the items answered correctly on the pre-test when compared to a count of those same items on the post-test will show a drop from what it was on the pre-test.

b. That an item count of items found to be correct on the delayed re-test when compared item for item with items found to be correct on the pre-test and post-test will tend to show that learning has taken place between the time of course completion and administration of the delayed re-test.

Analysis of the results yields the following conclusions:

1. The average score for the items answered correctly on the pre-test when compared to a count of the same items on the post-test showed a drop.
2. The average score for the items answered correctly on the pre-test when compared to a count on the same items on the re-test showed a drop.
3. Both score differences in 1) and 2) proved to be statistically significant.
4. The score for the items found to be correct on the re-test, that were not found to be correct on the post-test or pre-test, indicated that learning had taken place since course completion.
5. The differences between these scores proved to be significant.
6. Hypothesis a. 1), when used as described in the formula introduced in this study, causes a decrease in the percent retention. Hypothesis a. 2), when used as described in the formula introduced in this study causes an increase in the percent retention. Hypotheses b. causes a decrease in the percent retention. These differences, occurring in opposite directions, tend to balance each other out.
7. The experimental percent retention mean is less than the "conventional" percent retention mean. This occurs for both the February and September groups. The experimental percent retention mean proved not to be statistically significant from the "conventional" percent retention mean.

It is the author's feeling that mean scores for percent retention will always be smaller when calculated by the experimental method described in this study. If this proves to be so, then one-tailed tests of significance may show that a significant difference does exist.

REFERENCES

Anastasi, A. "Further Studies on the Memory Factor," *Archives Psychology*, No. 142 (1932), p. 60.

Ansbacher, H. L. "On the Permanence of College Learning," *Journal of Educational Psychology*, XXXI (1940), 622-626.

Ballard, P. B. "Obliviscence and Reminiscence," *British Journal of Psychology*, *Monograph Supplement No. 1* (1913), 82 pp.

Cedarstrom, J. A. "Retention of Information Gained in Courses in College Zoology," *Journal of Genetic Psychology*, XXXVIII (1930), 516-520.

Damrin, Dora E. "A Comparative Study of Information Derived from a Diagnostic Problem-Solving Test by Logical and Factorial Methods of Scoring," University of Illinois Doctoral Dissertation, *University Microfilms Publication, No. 4428* (1952), p. 6.

Doppelt, Jerome E. "The Correction for Guessing," *Test Service Bulletin*, the Psychological Corporation, No. 46, January, 1954.

Ebbinghaus, H. *Memory*, translated by H. A. Ruger and C. E. Bussenius, Teachers College, Columbia University, 1913.

Greene, E. B. "The Retention of Information Learned in College Courses," *Journal of Educational Research*, XXIV (1931), 262-273.

Johnson, P. O. "Permanence of Learning in Elementary Botany," *Journal of Educational Psychology*, No. 21 (1930), pp. 37-48.

Kennelly, T. W. "The Role of Similarity in Retroactive Inhibition," *Archives of Psychology*, No. 260 (1941), 56 pp.

Luh, C. W. "The Conditions of Retention," *Psychological Monographs*, XXXI (1922).

McGeoch, V. A. and McKinney, F. "The Character and Extent of Transfer In Retroactive Inhibition: Disparate Serial Lists," American Journal of Psychology, XLVII (1935), 409-423.

Odell, Charles W. A First Course In Educational Statistics, William C. Brown Co., Dubuque, Iowa, 172 pp.

Powers, S. R. "A Diagnostic Study of the Subject Matter of High School Chemistry," Teachers College Contributions to Education, No. 149 (1924), p. 84.

Robinson, E. S. and Heron, W. T. "Results of Variations in Length of Memorized Material," Journal of Experimental Psychology, V (1922) 428-448.

Tyler, R. W. "Wnat High School Pupils Forget," Educational Research Bulletin No. 9, Ohio State University, 1930, pp. 490-492.

Watson, I. R. "An Experimental Study of the Permanence of Course Material in Introductory Psychology," Archives of Psychology, No. 225, (1938), 65 pp.

Welborn, E. L. and English, H. "Logical Memory and Retention: A General Review of Experiments with Meaningful Verbal Materials," Psychology Bulletins, Nos. 1-20 (1937), 34 pp.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 2, December 1960)

PERSONALITY PATTERNS OF PROSPECTIVE TEACHERS

ROBERT F. PECK
University of Texas

THE PERSONALITIES of teachers have long been known to influence the effectiveness of their teaching (3, 5, 10, 12, 14). There have been at least two knotty problems, however, in establishing the exact nature of the complex relationships between personality and teaching effectiveness. First, there have not been very effective measures for those aspects of personality which are especially pertinent to teaching performance. Many early studies relied on existing "personality" questionnaires of a self-report kind. These were seldom designed to measure characteristics central to the teaching process; or, if they purported to be, they were prey to the reluctance of most respondents to appraise themselves frankly and accurately (2, 4). The other problem, of course, is the great difficulty in obtaining criterion measures of personality which are themselves valid and meaningful (11).

The present paper reports an exploratory study aimed at improving the definition of teaching-relevant personality characteristics, as these characteristics appear and are measurable in projective data. The findings may suggest new measures of these personality characteristics, in more quantitative form. The findings from this study of prospective teachers also suggest some rather disquieting features about the kinds of young people who are today preparing to teach. If later investigation verifies these data, a much increased guidance and personal counseling program would seem indicated as a necessary part of the professional preparation of at least half of our future teachers.

The present report can only be taken provisionally, for in line with Barr's observation that "there is nothing but trouble ahead in all research (or) it wouldn't be research" (4), despite an extra year of work following the study reported here, it proved impossible to both locate in teaching and have assessed (by supervisors) enough graduates from this sample of 69 college students to provide a criterion derived from on-the-job performance. There was validity evidence from comparable personality studies by the writer, as well as parallel data on the subjects in the form of self-appraisal instruments for the measurement of personal adjustment and maturity.

Procedure

A sentence completion instrument designed for use with college students, (Peck Sentence Completion, Form 2-B) was administered to a random sample of 69 junior and senior women majoring in elementary education at the University of Texas or at a state teachers college. A team of three --Kathleen Keen[2], Aubrey Roden[3], and the writer --jointly analyzed the protocols and rated the subjects on a nine-point scale of general "Teaching Potential."

Since the personality analysis was done in conference style, to maximize accuracy and penetration, the appropriate measure of reliability was a re-rating of the 69 cases, some months later. At that time, rate-rerate reliability was found to be .88.

The Teaching Potential rating required simultaneous weighing of two factors: the subject's ability to perceive, organize and communicate information in an objective, orderly manner ("academic" teaching ability), and her probable influence for good or ill on the mental health of her pupils. After all cases had been finally re-rated, the research team examined the subjects who were grouped at each scale point and constructed a qualitative description of salient personality characteristics of each group on the scale.

The scores on the Teaching Potential scale were then correlated with scores on four other personality measures which had been administered at the same time as the sentence completion. These included the Lefkowitz Rigidity Scale, the Worchel Self-Activity Inventory, the McCandless Anxiety Scale, and a Mental Health Q-Sort developed by the writer with the aid of McGuire and others.

---

* Footnotes will be found at the end of the article.

Results and Discussion

The correlations of the Teaching Potential rating with the other instruments are as follows: Rigidity Scale, -.42 (significant at .01 level); SAI total score (the weighted number of problems checked as self-descriptive) -.20; SAI self-ideal discrepancy score, -.33 (significant at .01 level); McCandless Anxiety Scale, .05; Mental Health Q-Sort, "maturity" score, .28 (significant at .05 level).

Self-descriptive measures such as these have uncertain validity themselves. Moreover, these are not all designed in such a way that one would reasonably expect a simple, linear relationship with such an overall rating as the Teaching Potential score. However, there is a significant degree of agreement with all but the McCandless instrument, in the expected direction. The T-P ratings correlate negatively with self-ascribed rigidity and with discrepancy between perceived self and ideal (desired) self. The T-P ratings correlate positively with the "maturity" score on the Q-Sort. Thus, so far as this evidence goes, it suggests that the Teaching Potential ratings tended to be valid. Perhaps this is about as much correspondence as one could expect when comparing judges' appraisals with self-appraisals in this area that is so loaded with the social desirability effect.

Ideally, depth appraisals by other trained judges, using other kinds of data, would be the desirable criterion for testing the validity of the personality descriptions. Failing this--primarily for reasons of research economics in the present instance--a next-best measure is to establish the credibility, or validity, of the judges (not just of the type of data used). Evidence from some analogous studies shows validity coefficients of this kind, of .65 to .83 (7, 8, 9).

The subjects grouped together at each of the nine points on the scale ("one" was the lowest score) were analyzed to determine what characteristics typified that group and differentiated it from the groups above and below it.

Scale Point One (Four subjects rated here). Two sub-types appear here: (a) A pattern of severe personality disorganization, expressed in acutely conflicting attitudes, very unstable emotionality and mental confusion. This disorganization is so manifest and strong that it appears likely to be very actively upsetting and confusing to pupils if these girls were to teach. The extreme example is a girl who seems unmistakably to be a full-blown, though very quiet, schizophrenic whose psychotic confusion apparently had not been evidenced dramatically enough to call official attention to her condition.

(b) A pattern of extreme hostility, openly and explosively acted out on frequent occasions. Perhaps the flavor of this may be conveyed by quoting the all too typical response of a young divorced woman who wrote "Most people don't know that I own a brat." And this was not the strongest of her numerous expressions of bitterness and hostility. In all cases in this pattern, as might be inferred from the lack of self-control, both the affect-states and the ego-functions of these girls were markedly childlike and unorganized. They were rated very low because it appeared almost certain that they would actively hurt, antagonize and deeply upset pupils in a way destructive both of mental health and of efficient learning of school subjects.

Scale Point Two (Eight subjects rated here). These girls show some active mental confusion, but they manage to "make sense" most of the time. They also reveal intense feelings of hostility; but unlike the "Ones," they control it and do not act it out except for sporadic temper outbursts when immediate stresses become too great for their repressive and suppressive system of restraints. They are, nonetheless, acutely and frankly unhappy in a deep, all-pervasive way. They are rated higher than the "Ones" chiefly because they appear less likely to assert an actively destructive or severely confusing influence on pupils. They would still be somewhat unpredictable, upsetting, and apt to spread their own strong unhappiness to a degree that is unpleasant even if it is not acutely disruptive.

Scale Point Three (Twelve subjects rated here). The girls in this group are actively unsatisfied with life; not painfully unhappy, as are the Ones and Twos, but restlessly discontented. They seem not only to have no clear, personal goals in life, but to lack "know-how" about ways of identifying activities that would be enjoyable, and ways of initiating such activities for themselves. What makes them passable performers, it seems, is their passive, if distinctly unenthused, willingness to accede to official demands which a directive system (such as a school, or their parents) may make of them. They would "walk through" the role of teacher, it seems likely, not finding much personal meaning or enjoyment in it, but also not exerting any actively upsetting influence.

These are girls, it seems, who could be hired to teach in almost any school, without anyone raising an eyebrow. There is nothing "wrong" with them, in an active sense. On the other hand, they would seem likely to be a drag on the school program and on their supervisors, reqiring more attention, more effortful pressure and more continuous guidance than it is easy or desirable to give. Lacking such stimulation, they appear likely to do a decidedly uninspired, unenthused, mechanical job of teaching, at best, and they are likely to leave a number of necessary tasks undone or half-done if not firmly and steadily led.

Scale Point Four (Fourteen subjects rated here). These girls, too, show a restless dissatisfaction with their lives, and an unfocused but strong longing for more "point" to life. Their dominant mood-tone is one of moderate unhappiness or discontent, but it is much more attenuated than is true of the lower groups. In addition, these girls show an active conscientiousness. They are not very well organized in the way they think and act, but given some helpful guidelines to follow, they go at tasks with a measured degree of active interest and purposefulness.

It may be that, unlike any of the preceding groups, these girls could rather quickly and readily be trained to be effective teachers and purposeful, reasonably well organized people in general, if they are given active, interested, personal, rather intensive on-the-job leadership and training for some two or three years. Unlike the lower groups, they have some degree of positive interest in life, already. It is quite vague, unharnessed and unpracticed; it is also overshadowed by somewhat more pervasive negative feelings about life and self; but it is there as a present motivation which could be brought out and developed.

It seems clear, however, that it would require personally interested, individualized leadership to bring out this potentiality. No formal system of training or "leadership", however well organized in theory, would "reach" these girls, if it lacks the personal element of a leader or guide who feels a genuine, human interest in each girl as an individual. This is not intended as a Peale-ish homily, on rosy "general principles." Rather it stems from the specific lacks and needs these particular girls appear to demonstrate. Their lack is not a lack of formal training: they have already had sixteen years of schooling, including teacher training. Their need, and their lack, appear to go much further back into their family years, to be much more intensely personal, and to concern profound emotional needs more than a need for conscious intellectual training alone.

So far, all the intellectual training they have received has not "taken" very deeply or usefully, not because of deficient intelligence, but apparently because of their continuing preoccupation with vaguely perceived but intensely felt needs for relatedness-security. It might be noted that "good leadership" for them would have to be patiently understanding, too, because most of these girls would humanly resent and defensively resist a "critical" pointing out of their very real personal problems. Finally, it might be that effective guidance could rather economically mobilize and put to work all the intellectual training which they have gathered and which at present must be lying latent in their minds, unorganized and largely unused.

Scale Point Five (Eight subjects rated here). These girls might be summed up as quiet, responsible, efficient conformers. They show very little active, personal initiative but once they are given a role to fulfill, they perform in a sensible, orderly way. They show no particularly active enthusiasm for life or for teaching, but neither do they experience much unhappiness or marked dissatisfaction. They appear to be quite stably what they are: not at all inspired or inspiring, but quietly competent. If they would not arouse strong learning motivation in pupils by their own interested example, they would nonetheless carry out their assigned teaching duties in a calm, relatively cool, impersonal way. They do not demonstrate any active, special, personal interest in teaching, whether as an avenue for self-realization, as a way of establishing gratifying human relationships with children, or in any other way. It is a job; if they take it, they will do it as prescribed; and that seems to be that.

Scale Point Six (Four subjects rated here). As one moves up the scale, this is the first group which shows a degree of active, independent purposefulness. These girls are also the first group to show a predominantly pleasant, good-natured reaction to life. They have some personal goals which they can actively pursue when circumstances actively permit and encourage it. They are not especially well organized nor extremely purposeful in the way they act. Much of the time, they cheerfully go along with the social and professional currents around them, in a spontaneous but rather unthinking way. They can muster some personal initiative, however, if there is something they want very much and no one else is providing it for them. In their own way, they appear to have a relatively stable, reasonably satisfied way of fitting in with life as they know it. They are not apt to exert much sustained effort to understand a particular child (or adult) in order to meet that other person's special need. But their customary good nature has a pleasant effect on others' feelings in a general way. Furthermore, these girls carry out their duties, once assigned. with some degree of active thought as to the purpose of what they are doing; or at least about how well it is or isn't working. They do not proceed in a purely mechanical or ritualized manner. They are neither much interested, at base, in being expert intellectualizers, nor prone to exercise much imagination about new or better ways to teach some topic. Nevertheless, to a mild degree they probably would "adapt the subject matter to the child," largely because they are most comfortable when everyone around them is comfortable--and a school pupil is apt to be most "comfortable" when what he has to do makes sense to him.

Scale Point Seven (Five subjects rated here). This group presents the first instance (ascending the scale) of a reasonably well-defined, firm ego structure. These girls show a distinctive, if in most cases still-emerging, individuality. They are working at defining goals in life which they look toward with personal interest. Their general feeling about their past is not one of unalloyed pleasantness; some parts were and are painful. However, instead of giving in to that, they are actively striving to find positive rewards in the present and future; and they feel they are finding and will find them. They are somewhat restless, experience some marked uncertainties and inner conflicts about what they "really want to be, and do"; but on the whole they are actively in the process of becoming masters of their feelings, their actions and their own lives, rather than being the passive pawns of circumstances and external pressures.

They are not quite as stable, well-organized and vigorously purposeful, especially in the direction of a teaching career, as the Eights and Nines, though notable individual differences exist. Two girls are powerfully motivated toward social mobility, and well equipped to achieve it, it appears. One of them is distinctly not going to teach, or not more than a year or two, it seems certain. She will probably either marry a rising young man or find an individually expressive career in journalism or an allied field. The other "mobile" girl may become a career teacher, if children of her own do not intervene; but here her almost too-powerful manipulative drive appears to somewhat reduce the healthiness of her impact on pupils from an otherwise excellent level of effectiveness.

One of the other girls in the group is frankly far more interested in marriage and a family of her own than in a teaching career. If she did teach, she would be a lively, stimulating person, but off-hand enough in her attitude to reduce her effectiveness from top level. The other two girls are more shy, more interested in teaching, though also more unsure of themselves and somewhat hesitant to take firm stands. Underneath, though, they have a growing firmness which could well surprise people who "know them well;" or think they do. These girls are extremely sincere, as indeed are all the girls at this level and higher.

Thus, for somewhat varied reasons, the girls in this group show good capacities and positive attitudes toward life. Some are here, rather than higher, because of a pronounced lack of interest in teaching as an occupation; some, because of moderate personal problems which keep them, as yet, from full, unhampered application of their basically solid, constructive aptitudes and attitudes.

Scale Point Eight (Nine subjects rated here). These girls are, to a woman, extremely sensible, well organized and clear-sightedly purposeful. By contrast with most members of the preceding groups, they are all characterized by a high level of available and applied energy. They go at life with considerable zest, if not with unmixed pleasure, then with firm determination to make the best of it. They are thoroughly realistic. They also find life much more satisfying than not. They have firm self-confidence, on the whole, and take responsibility for the way their lives work out. If things go wrong, they may dislike some other person who may have been partly responsible; but instead of moping or feeling prolonged resentment about it, they figure out ways to create new satisfactions for themselves.

These subjects were rated below the Nines, partly for varied, individual reasons, partly on general grounds of not quite as high a degree of creative, applied intelligence, not quite as complete and mature autonomy, and/or not quite as strong and specific a dedication to teaching as a major personal interest. Apart from these differences of degree, there are no marked differences between the Eights and Nines.

Scale Point Nine (Five subjects rated here). These women (whatever their age, psychologically they appear too mature to be termed "girls") possess outstanding degrees of several characteristics, all bound together into a unified pattern. While thoroughly human, in that they know what it is to experience conflict, disappointment and frustration, they are both solidly in command of their lives and very wholeheartedly capable of dedicating themselves to teaching as a major, deeply satisfying pursuit. There are many notable differences in personality among them. What they share seems to be these things: high intelligence very actively, resourcefully and ingeniously applied; a clear, long-range perspective, both on specific tasks and on life as a whole; an extremely vigorous, unreserved drive to accomplish (not, it must be said, a neurotically over-determined drive); a very sensible, down-to-earth practicality and realism (they can see what can be done and do it, what cannot be done, and turn to other pursuits which are feasible); and they have a high degree of psychological autonomy. This last feature does not mean indifference to other people. Quite the contrary, in one way or another these women feel cheerfully responsible to treat others in a decently respectful, constructive way. Moreover, just as they make strong emotional investments in their chosen avenues of achievement, when they build relationships with other people, these people become deeply important to them emotionally. These women do not show either the hysterical "gaiety" or the impulsive, half-projective over-investment that many of the girls at levels Three and Four (or even Two) exhibit. Rather, their heads and their hearts, it might be said, work together. The result is that they can care deeply about other people--or, in one case, about her responsibilities to other people, at least--and at the same time they can maintain objective, accurate judgment.

Finally, these women were rated highest on teaching potentiality for the rather self-evident reason that, in one way or another, they actively choose teaching as a personally sought career. This, added to their aptitudes, maturity and purposeful vigor, seems reason to expect them to perform extremely well, both as academic instructors and as influential examples of thorough maturity and emotional health.

Implications of the Teaching-Potential Distribution

Without further validation, the distribution of ratings assigned to the 69 subjects cannot be uncritically accepted as a statement of fact. Still, if the ratings even roughly approximate the true picture, there would be a number of practical implications of considerable importance. The overall distribution is as follows; and it might be noted that these subjects presumably represent the "cream of the crop" of the nation's youth, both intellectually and in social acceptability.

Scale Point:	1	2	3	4	5	6	7	8	9
No. of Cases:	4	8	12	14	8	4	5	9	5

This is skewed toward the high end. The distribution is distinctly a bi-modal one, however; and if the psychological meaning of the scale points is considered, the appropriate break-point in terms of desirable personal qualities for teaching might be drawn between points 5 and 6. In that case, it appears that only one-third of the total sample show the traits one would like to have in all teachers. Two-thirds range from a quietly uninspired, somewhat rigidly conformist pattern (Point 5) through a large group of aimlessly discontented girls (4 and 3), down to a group of acutely unhappy, confused, actively upsetting girls who look quite undesirable as teachers of children(2 and 1).

Even using a conservative criterion which calls only groups 1-4 "below average," 38 of the 69 subjects show characteristics which are scarcely desirable in a teacher. They are: discontentedly unhappy, in general, increasing to bitter hostility at the lowest levels; lacking meaningful personal goals and lacking effective skill at setting and pursuing such goals; restless, aimless and unorganized in their thinking, ranging downward to extreme confusion at the lowest level; and they are mistrustfully lacking in any spontaneous friendliness toward other people a great deal of the time, ranging down to open, destructive hostility.

The 12 girls at levels one and two probably should not enter teaching. They are too confused and hostile to do other than disturb children and actively interfere with their learning. These subjects could appropriately be termed severely neurotic; and they include one who shows every sign of being frankly psychotic. In a sense, these girls should not present an insuperable practical problem. Barring intensive psychotherapy, they should be excluded from teaching. This is an "administrator's" point of view, of course.

In actuality, it seems a major indictment of the depersonalized ignorance about individuals in colleges of education that girls like these should have been permitted, even encouraged, to go through the entire program of teacher preparation. Even a crude screening on grounds of personal attitudes and adjustment could identify most of these girls before they ever entered a teacher-training program. Thereafter, if either the need for teachers or a desire to help these girls indicated, some measure of adjustment guidance could conceivably be provided. Without it, they remain extremely disturbed in their personal lives and almost certainly unfit to teach. In this sample, they constitute a disquieting 17%; and they are even more likely to go into teaching and stay in it than are the girls at the highest three levels on the scale. The "high" girls have very attractive qualities, are interested in marriage, and many seem likely to marry out of teaching permanently. These "low" girls, on the other hand, are scarcely attractive prospects for marriage in their personalities, even though they too would like to marry. It appears not unlikely that many of them will take the opportunity to make a lifetime career of teaching. To be blunt about it, they can probably get teaching jobs when no one else or no other kind of organization would take them, or retain them long.

In many ways, though, the most challenging problem is presented by the 26 girls at levels 3 and 4 on the T-P scale. They constitute over one-third of the combined sample. As the descriptions indicate, they are not completely lacking in potential ability to teach adequately, but they are far from what any school administrator would knowingly pick for his teaching staff, if he had any free choice. Yet these girls appear significantly unlike the ones at the lower levels. They are aware, however vaguely and unhappily, that their lives lack meaning and purpose. They would like to change this, if they only knew how. That they do not know how seems to reflect a learning deficiency rather than a neurotic over-learning of irrational reactions. They resemble children who have been taught, life-long, to do nothing "until Momma or Papa tells you," by parents who are too preoccupied or too blindly imitative of ritualized conventions to do much positive "telling." In this respect, these girls resemble even those at level Five, to some degree. Their problem seems less one of distorted learning (neurotic or psychotic) than of an absence of learning. In particular, they lack the skills to conceive and initiate important, rewarding activities, and they also lack a confident sense that it would be right and proper to initiate such autonomous seeking for meaning and satisfaction in life.

The most apt phrase for these girls' problem may be "psychological emptiness." This phrase, at least, seems to convey the difference between the conflicted, neurotic states exemplified by the subjects on levels One and Two, and the relatively unconflicted but aimless, childishly "helpless" state which these girls at the next higher levels are experiencing. In many ways, this state resembles what David Riesman has described in The Lonely Crowd--especially his "Outer-directed" type of personality.

These girls, too, are most likely to fall back on a teaching career for want of any personally defined goals which have an active appeal. For most of them, teaching presents the only area of "guaranteed" security and orderly purpose which they have found or can conceive of finding. They are not lacking in intelligence: at least they have almost completed college. If, moreover, they are among those most likely to seek, get and keep teaching jobs, it would seem highly desirable to give them much effective personal guidance during their college years, and effective on-the-job guidance. At present, it is an unwelcome but recognized fact that neither adequate college guidance nor attentive, individualized in-service guidance is being provided for people like this, except in extremely rare instances.

The very fact that so many "wishy-washy," socially and intellectually inept students could get this far in professional training seems uncomfortably clear evidence that there is a profound lack of knowledge about individual students on the part of their professional mentors, the faculties of their colleges. For if these girls were recognized for what they are, and still nothing was done, it would not speak very well for the professional integrity or the standards of their educational mentors.

Perhaps, though, this is a most unfair interpretation. With the seriously low level of budgetary support for these colleges of education, the undergraduate student-teacher ratio is known to be far too high to permit any but an almost overly zealous professor to get to know his or her students at all well, as individual people. Furthermore, the extremely scanty funds and staff available for guidance make intensive guidance almost impossible, in all but a very few cases. Indeed, in the words of one university administrator, over 90% of what little guidance is available is spent on students who have passed the point of no return through academic failure. In brief, a practical person might conclude that teaching positions will inevitably continue to be manned by aimless, over-dependent, intellectually inept, discontentedly unhappy women in one-third to one-half of all instances. In that case, perhaps we should stop preaching empty platitudes about "good teaching." Indeed, we might well save much time now largely wasted on such trainees, in the illusion that we are "training their minds." If these girls are any evidence, we aren't even "reaching" them in any personally meaningful way.

Yet such a pessimistic resignation is not only repugnant; it is actually contradicted by certain evidence in the present data. For just these girls, in particular, school seems to have been the one area of life where they have predictably found some measure of meaningful order, kindly, encouraging treatment by some teachers, and a modicum of self-respect through the learning of some purposeful skills. Clearly, this has not been enough to compensate for earlier, family-based lacks, nor to bring them farther along than they now stand in their personal maturation; but just as clearly, it suggests that schools and colleges have an already-established foothold with these girls. Whatever extra effort could possibly be invested, perhaps in the general kind of guidance suggested for the Scale Point Four subjects above, such effort appears to have a good chance of paying off in the form of better organized, more enthused, interested teachers (13).

Needless to say, the one-third of the sample at levels Six through Nine require little more than active encouragement to be their healthy, good-natured, intelligently purposeful selves. Conventional professional training seems well suited for the preparation of such teachers-to-be. As far as concerns future research, and educational policy, people like these might well serve as the models for criterion-development in the defining of "effective teaching." Certainly, they represent many diverse patterns of interest and behavior-style. There is no one pattern of "the good teacher," here. Some are warmly friendly, a few are firmly impersonal. Some are gaily outgoing, some are calmly reserved. Some put their major emphasis on intellectual clarity and skill; some put it on building friendly, encouraging personal relationships with colleagues and pupils. (None chooses one of these goals to the exclusion of the other.)

For such reasons of personal difference, differential prognoses were made in many cases, concerning the grade-level or subject-matter area where a particular girl seemed likely to function with peak effectiveness. A warm but shy, slightly uncertain girl could be seen as a highly effective primary teacher. To put the same girl to teach in many junior high schools might be tantamount to "throwing her to the wolves." Conversely, a hyper-energetic girl who loves and must have continuous freedom for physical action, and who prizes athletic skills, might be a comparative misfit as a classroom teacher, however intelligent and healthy she is. The same girl might prove outstanding as a physical education teacher, and might also develop into a wise, welcome personal counselor for many school girls, as long as she had a position that gave her full outlet for her sheer high level of physical energy.

This kind of attempt to match the particular person with a particular job is becoming a rather accepted procedure in industry. It is only its application to teacher-training and placement which is relatively new and untried.

Summary and Prospect

Personality analyses were made of 69 undergraduate women majoring in elementary education. The available validity evidence moderately tended to confirm an overall rating on "teaching potential," based on personality analyses from sentence completion data. Salient characteristics of the subjects at each level on the T-P scale were described, and implications for teacher education were discussed.

If these personality characteristics can be measured by additional methods, while converting the qualitative analysis of the projective data into quantitative measures, a means might be developed for predicting teacher effectiveness with increased accuracy. Such an undertaking is planned as part of the new five-year research program of the Mental Health Demonstration Center at the University of Texas, supported by a grant from The National Institute of Mental Health. Validation is planned against both expert personality appraisal from independent data and observational data on teaching performance in the years after college.

FOOTNOTES

1. This research was part of a program supported by funds from the Hogg Foundation of Mental Health.
2. Now at the Menninger Foundation, Topeka, Kansas.
3. Now at the University of California, Berkeley, California.

REFERENCES

1. Alexander, T. "Prediction of Teacher-Pupil Interaction with a Projective Test," Journal of Clinical Psychology, July, 1950, 6, 273-276.
2. Barr, A. S. The Measurement of Teaching Ability, Madison, Wisconsin: Dembar Publications, Inc., 1945.
3. Barr, A. S. et al. "Measurement and Prediction of Teacher Efficiency," Review of Educational Research, June, 1955, 25, 261-269.
4. Barr, A. S. "Problems Associated With the Measurement and Prediction of Teacher Success," Journal of Educational Research, May, 1958, 51, 695-699.
5. Hunt, J. T. "School Personnel and Mental Health," Review of Educational Research, December, 1956, 26, 502-521.
6. McGuire, C. and Peck, R. F. Assessment of Personal Effectiveness and Potential Teaching Competence. Paper presented at American Educational Research Association, February 23, 1958.
7. Peck, R. F. "Measuring the Mental Health of Normal Adults," Genetic Psychology Monographs, 1959, 60, 197-255.
8. Peck, R. F. and Parsons, J. "Personality and Work Output," Personnel Psychology, Spring, 1956, 9, 49-79.
9. Peck, R. F. and Thompson, J. "The Use of Individual Assessments in a Management Development Program," Journal of Personnel Administration and Industrial Relations, April, 1954, 1, 79-98.
10. Ryans, D. G. "The Investigation of Teacher Characteristics," Educational Record, October, 1953, 34, 371-396.
11. Ryans, D. G. "Notes on the Criterion Problem in Research, With Special Reference to the Study of Teacher Characteristics," Journal of Genetic Psychology, 1957, 91, 33-61.
12. Ryans, D. G. "A Study of Criterion Data (A Factor Analysis of Teacher Behaviors in the Elementary School)" Educational and Psychological Measurement, Autumn, 1952, 12, 333-344.
13. Simpson, R. H.; Gaier, E. L. and Jones, S. "A Study of Resourcefulness in Attacking Professional Problems," School Review, December, 1952, 60, 535-540.
14. Symonds, P. M. "Characteristics of the Effective Teacher Based on Pupil Evaluations," Journal of Experimental Education, June, 1955, 23, 289-310.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 2, December 1960)

FACTOR ANALYSES OF THREE TESTS OF CRITICAL THINKING*

VELMA I. RUST
University of Illinois

FOR SEVERAL decades measurement of critical thinking skills and abilities has presented a major difficulty because educators have not agreed on the kind of measurement which should be used. There has even been a lack of agreement regarding a definition for critical thinking. The problem of measurement is the focus of this study. In a search for common elements in the operational definitions of critical thinking used by different researchers, three tests of critical thinking were analyzed.

This research arose from the findings of the Illinois Curriculum Program Committee relative to its Critical Thinking Project (9). When the Project Committee and the cooperating teachers wanted to evaluate their attempt to teach the principles of critical thinking, the following three tests were selected for the purpose: the Watson-Glaser Critical Thinking Appraisal, Form Bm; A Test of Critical Thinking, Form G, prepared by the American Council on Education; and A Test on Principles of Critical Thinking, Form Fl. 5, prepared by the Project staff (hereafter called the Watson-Glaser, ACE, and Principles tests respectively).

Low intercorrelations were obtained between total scores on any two of these tests. These low intercorrelations led to the inference that there is little relationship between the knowledge of the principles of critical thinking and the ability to use these principles. The Project Committee consequently felt the need for more information regarding the internal structure of the three critical thinking tests.

The technique of factor analysis was used for the purpose of determining whether the tests actually do measure what they are supposed to measure according to the a priori reasoning of the test makers. The research hypothesis tested was that factor analyses of these three tests of critical thinking will yield some common factors among the items in these three tests.

Objectives

Specifically, the objectives of the study were threefold. The first of these was to study each test individually by factor analysis in order to check the a priori reasoning of the makers of the test regarding appropriate grouping of items. It was thought that such studies would give more insight into the internal structure of the tests themselves. It was also hoped they would yield some clues regarding the correctness of the judgment of the Illinois Critical Thinking Project Committee that several items have no place in tests of critical thinking and might well be deleted.

A second objective was to plan a factor analysis of a composite test made up of selected items from all three tests for the purpose of ascertaining whether or not the three tests measure the same skills and abilities. Moreover, it was reasoned that with the information gathered from the four factor analyses, viz., the Watson-Glaser Critical Thinking Appraisal, the ACE Test of Critical Thinking, the Principles of Critical Thinking test, and a composite test, it should be possible either to improve the Test on Principles prepared by the Project Committee or to enable a new test to be developed.

The improvement of the Principles test became the third and most important objective. In other words, an effort was made to determine whether or not the Principles test actually tested what it was supposed to test according to its makers.

Procedure

From the group of about three thousand students involved in the original Critical Thinking Project, 949 students were selected, namely, those who took the Watson-Glaser, the ACE, and the Principles tests. The answer sheets for these students for all three tests were carefully scrutinized. If any item on any test was omitted by a student, that student's scores were discarded. In this way the number of subjects was reduced to 587. A side investigation later completed showed that this sampling technique had not created bias with respect to intelligence test scores or to scores on the achievement tests administered by the Project staff.

---

* This report is a summary of a major piece of research completed in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Education in the Graduate College of the University of Illinois. For assistance with the preparation of the dissertation, the author is deeply indebted to the members of her thesis committee, namely, Dr. Kenneth B. Henderson, Chairman; Dr. R. Stewart Jones, Dr. D. L. Burkholder, Dr. Franz E. Hohn, Dr. Philip J. Runkel, and Dr. Henry F. Kaiser.

Before carrying out the factor analysis, each test and its manual were perused carefully for the purpose of discovering what the test makers had in mind when they prepared their test, that is, what they believed critical thinking involved and what aspects of it they hoped to test. Inter-item correlation matrices were then obtained separately for the three tests. Phi coefficients were used since the data were dichotomous.

The centroid method of factoring was employed. Since the purpose of a factor analysis here was to analyze the common factors that might explain the intercorrelations, communalities, instead of unity, should have been written in the diagonals. However, at the risk of increasing the rank of the correlation matrix and of forcing the unique variance into the common factor matrix, it was decided to use unity in the diagonal cells of all matrices. This decision was justified on the following grounds: (a) Because of the large size of the correlation matrices, the factorial results would not be seriously affected by coarse initial estimates for the diagonal cells; (b) To use any known formula to compute better estimates would have multiplied the required computer time by at least six or seven times, and any slight increase in accuracy gained at this stage did not seem to warrant such expenditure of time.

The factor analysis procedure yielded only one weak general factor for each of the three tests. It was, moreover, practically impossible to estimate how many factors should be sought. The gradations in the amount of variance accounted for were so slight that one could have gone on factoring indefinitely.

At this point it was decided nothing could be gained by attempting to rotate to oblique simple structure, or by factor analysis of a composite test. It was believed that since no strong common factors within each of the tests separately were discovered, and since correlations between any two of the three tests were low, there was little likelihood that there would be found any strong factors common to all three tests. Instead, a final attempt to get additional clues regarding the internal structure of the tests was made by throwing on the oscilloscope of the Illiac plots of pairs of factors for the first eight centroid factors and photographing these. Since the results were not encouraging, no further use was made of the computer or the factor analysis procedure.

Findings

Table I shows what the makers of the three critical thinking tests were hoping to test. It will be noted that, to some extent, the different tests emphasize different aspects of critical thinking.

Compilations of frequencies of correct responses showed that the Watson-Glaser and ACE tests were easy for the group of students concerned. Mean item difficulties for the tests were: Watson-Glaser, 0.66; ACE, 0.67; and Principles, 0.51. On the whole, inter-item phi correlation coefficients were low. The higher correlations tended to cluster along the diagonal, thus lending support to the a priori reasoning of the test makers regarding grouping of items. The size of phi correlation coefficients between pairs of items of the Watson-Glaser test ranged from approximately -0.325 to approximately +0.675. There were only six cases of a phi coefficient higher than +0.375, and the highest correlation coefficient, which was +0.626, was between items 38 and 39 which the test makers grouped under the subtest labeled "Deduction." In the case of the ACE test, phi coefficients ranged from approximately -0.125 to approximately +0.725, and for the Principles test from approximately -0.175 to approximately +0.375.

Loadings of items on the first centroid factor for the Watson-Glaser Critical Thinking Appraisal varied from about -0.15 to about +0.54. For the ACE test, loadings of items on the first centroid factor varied from about +0.15 to +0.54; whereas for the Principles test, loadings of items on its first centroid factor varied from about -0.20 to about +0.49. Table II shows the variance accounted for by the first centroid factors.

In general, items having highest loadings on the first centroid factor for each test came from so many different subtests and therefore represented so many different abilities that it would have been difficult to find a label for these common factors. The a priori reasoning of the test makers, however, which led to the grouping of as many as seven items, namely, 38, 39, 42, 48, 53, 55, and 57, in the Watson-Glaser test was confirmed.

Since each of these common factors accounted for only a small portion of the variance, it was of little value in explaining the intercorrelations of test items within the test. Thus very little parsimony was obtained insofar as describing the internal structure of each test in simple terms is concerned.

Conclusions

All these findings indicate that only in rare in-

TABLE II

THE VARIANCE ACCOUNTED FOR BY THE FIRST CENTROID FACTORS FOR EACH OF THE THREE CRITICAL THINKING TESTS

Factor	Variance accounted for (in percent)			Cumulative variance (in percent)		
	Watson-Glaser	ACE	Principles	Watson-Glaser	ACE	Principles
1	7.076	14.782	7.610	7.076	14.782	7.610
2	3.148	3.601	2.890	10.224	18.383	10.500
3	2.470	3.527	2.877	12.694	21.910	13.377
4	2.171	3.347	2.768	14.865	25.257	16.145
5	2.022	2.928	2.730	16.887	28.185	18.875
.	.....	....	.....	....	....	....
12	1.539	2.091	2.319	28.694	45.222	35.876

stances was the a priori reasoning of the test makers regarding grouping of items confirmed. This suggests that all items within a subtest do not measure the same skills or abilities, and therefore do not measure the skills or abilities they were intended to measure according to subtest titles and/or the manual accompanying the test. One might, then, conclude that the tests are poor tests and should be replaced. The low intercorrelations of test items and of total scores on the three tests lend support to this conclusion.

On the other hand, perhaps the tests are good tests, and the critical thinking process involves a large number of unique abilities and items of knowledge. The small portions of the variance which the first, and strongest, centroid factors accounted for, serve as evidence to support this conclusion.

The author feels that the evidence is not conclusive that the tests are good, nor is it conclusive that the tests are poor. Hence the author takes the position that the tests could be improved. More might have been learned about the internal structure of the tests if: (a) a different way of scoring the answers had been used (3), for example, weighting the responses or revising the keys (4); (b) a favorable testing situation had been ensured; (c) other subjects had been used, for instance, scores of subjects in the experimental group might be analyzed separately from those in the control group, or the sample used might be selected more carefully by casting out those whose responses on the Watson-Glaser test indicate the presence of a mental set (4); (d) control had been maximized when the Critical Thinking Project Committee set up the experiment; (e) the principal axis method of factoring had been used, and/or (f) a factor analysis had been made of the total scores for the subtests of each test instead of the individual item scores.

It is especially interesting to note that factor analysis of the items of three different tests purporting to measure critical thinking abilities yielded only one weak general factor each. One wonders what the implications of this result are. Does critical thinking ability not resemble other abilities in having a general factor? One asks this question because intelligence tests, tests of psychomotor and physical abilities, performance tests and tests of occupational abilities, all have yielded a thoroughly established general factor (within a given test) (10). It should be remembered, however, that most of the results of experimental work in this field have been based upon factor analysis of batteries of tests or subtests, not individual test items. Moreover, the finding here may be just a reflection of the low intercorrelation of test items, for a factor may be defined as a "construct which accounts for the objectively determined correlations between tests....", a category "for classifying mental or behavioural performances, rather than entities in the mind or nervous system." (10)

If one surveys research reports in the field of factor analysis of reasoning and intelligence tests (2, 5, 6, 7), one discovers the findings can be summarized as follows: (a) in all instances the researchers have been successful in isolating several first order factors; (b) in a few cases one or more strong second order factors have been extracted (8); (c) practically no information has been gleaned regarding any of the three critical thinking tests involved in the present study, except for one investigator's finding of low reliability for the Watson-Glaser test (1).

The factor analyses described in the present study yielded, on the contrary, no strong centroid factors. Therefore, further investigations regarding the characteristics of the subjects, test items, and critical thinking ability itself would seem needed. Two such worthwhile studies might be an investigation to determine whether a factor analysis of the scores on subtests yields the same results as a factor analysis of the scores on individual test items, and a factor analysis of a smaller sample selected in such a way as to control more variables.

REFERENCES

1. Bledsoe, J. C., "A Comparative Study of Values and Critical Thinking Skills of a Group of Educational Workers." Journal of Educational Psychology, 1955, 46, 408-417.
2. Blotzum, W. A., "A Factorial Study of the Reasoning and Closure Factors." Psychometrika, 1951, 16, 361-386.
3. Damrin, Dora E., A Comparative Study of Information Derived from a Diagnostic Problem-Solving Test by Logical and Factorial Methods of Scoring. Unpublished doctoral dissertation, University of Illinois, 1952.
4. Ennis, R. H., "An Appraisal of the Watson-Glaser Critical Thinking Appraisal." Journal of Educational Research, 1958, 52, 155-158.
5. Green, R. F.; Guilford, J. P.; Christensen, P. R., and Comrey, A. L., "A Factor-Analytic Study of Reasoning Abilities." Psychometrika, 1953, 18, 135-160.
6. Guilford, J. P.; Christensen, P. R.; Kettner, N. W.; Green, R. F., and Hertzka, A.F., "A Factor-Analytic Study of Navy Reasoning Tests With the Air Force Aircrew Classification Battery." Educational and Psychological Measurement, 1954, 14, 301-325.
7. Hertzka, A. F.; Guilford, J. P., Christensen, P. R., and Berger, R. M., "A Factor-Analytic Study of Evaluative Abilities." Educational and Psychological Measurement, 1954, 14, 581-597.
8. Matin, L., and Adkins, Dorothy C., "A Second-Order Factor Analysis of Reasoning A-

bilities." Psychometrika, 1954, 19, 71-78.

9. Smith, B. O.; Henderson, K. B.; Swift, L. F. Ennis, R. H.; Aschner, Mary J. and Payette, R. F., Teaching of Critical Thinking. Urbana, University of Illinois Press, in press.

10. Vernon, P. E. The Structure of Human Abilities. New York: John Wiley and Sons, 1950.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 2, December 1960)

PROSPECTIVE TEACHERS' ESTIMATES OF ADOLESCENTS' RESPONSES TO F SCALE ITEMS

JACK R. FRYMIER
Board of Public Instruction
Orlando, Florida

CAN PROSPECTIVE teachers accurately determine with which items of the F scale (a measure of anti-democratic potential) adolescents will tend to agree? Further, can they accurately determine with which items of the F scale adolescents will tend to disagree? If these prospective teachers do accurately predict adolescent responses, what will this mean? If they do not accurately predict their responses, what will this mean? In other words, if there are differences between their estimates and the way adolescents actually do respond, what will these differences mean?

Scodel and Mussen (1) reported that low F scale scorers were more accurate in estimating high F scale scorers responses to the measure than were high scorers in estimating the low scorers responses. Cohn (2) reported that a group of college students successfully faked responses to the F scale, implying that they could estimate with some degree of accuracy the responses of certain kinds of persons to the measure.

The Procedure

To answer the questions stated above, the author administered the F scale to 104 high school students in Michigan and Alabama and to 69 college students enrolled in a Teachers College of a Pennsylvania university.

Each high school student from Michigan was matched with one from Alabama on the bases of age, sex, and grade in school, making 52 pairs in all. The mean age of these two groups was 16.38 years, and each group had 22 girls and 30 boys.

The college group consisted of 38 boys and 31 girls enrolled in courses in the professional sequence leading to a Bachelor of Science degree in education. The mean age of this group was 22.50 years. All of the students but four intended to teach in secondary schools upon graduation. Of the four who did not intend to teach, three were preparing for the ministry and one for social work.

The responses of the high school students were handled in the following manner: Agreement or disagreement with individual items was recorded, i.e., any item which received a plus mark (+1+2 +3) was recorded as agreement, whereas any item which received a minus mark (-1-2-3) was recorded as disagreement. The percent of each group who agreed with each item was determined and the various items were then listed in rank order, according to agreement, and these rank orders were compared (3).

The ranking of the items for the two groups of high school students was not identical (there was not a perfect correlation for every item), although there was a large area of agreement. The author observed that although the proportion of the two groups who agreed with any particular item differed, in general both groups tended to accord approximately the same value to most of the various items as they related to the whole. That is, even though 75% of the Alabama group agreed with item 13 and only 56% of the Michigan group agreed, if the percent of agreement for all of the various items were given a rank order for both groups, both the Michigan and Alabama groups accorded this particular item approximately the same relative position, viz., thirteenth. There were some variations in this relative rank order between groups, but the general pattern seemed to be quite similar. In fact, only four items differed more than four positions between the two groups, whereas 23 items were accorded approximately the same value (i.e., differed four or less positions) in the overall hierarchy of values which seemed to be demonstrated by a complete ranking of all the items.

For purposes of this study, however, the responses of the adolescents from both groups combined was considered, and the final ranking of all the items was for all 104 high school students. The discussion above was designed to illustrate that although the respondents came from different cultural areas, the general pattern of agreement and disagreement was quite similar.

The responses of the college group were handled in the following manner: After responding to the F scale in the usual way (in order to insure familiarity with all items), they were directed to mark those three items with which they felt high school students would be most likely to agree. Further, they were then directed to mark those three items with which they felt high school students would be most likely to disagree.

Any item estimated as an agreement was given one plus point, and any item estimated as a disagreement received one minus point. The total value for each of the 27 items was determined arithmetically, and then all of the items were placed in a rank order according to the college students' estimated agreement through disagreement.

Finally, the actual rank order for the F scale items by the high school students was compared with the college students' estimate of high school students' rank order by inspection. Those items which differed less than four positions were considered as correctly estimated, whereas those which differed more than nine positions were considered as incorrectly estimated. That is, if the relative ranking of the college students' estimate of the high school group were less than four positions apart, the items were considered as having been correctly estimated; if they were more than nine positions apart, they were considered as incorrectly estimated.

The Results

The results of these procedures are recorded in Tables I and II. Table I shows the ranking of those eight items which the prospective teachers correctly estimated. These data seem to indicate that in respect to these items the prospective teachers correctly perceived adolescent values, as values are reflected in a hierarchy which is apparent in any ranking according to agreement.

Table II shows those items on which prospective teachers incorrectly estimated how high school students would agree or disagree. These data seem to indicate that this group of prospective teachers seriously erred in estimating adolescents' response to these particular twelve items of the F scale.

Discussion of Results

It seems that this particular group of prospective teachers recognized that adolescents appear to have faith in themselves to get things done, that science will not solve all problems, and that there are evidently times when high school youngsters want to be left alone, to have secrets, and to have certain things which are private and not known to other persons.

Further, this group of college students who plan to teach was correct when they thought high school students would not accept the idea that "familiarity breeds contempt," that persons must suffer to learn, or that "some people are born with an urge to jump from high places."

Finally, they also recognized that adolescents apparently are not generally agreed upon the idea that there will always be wars or that people must be careful in groups, since these high school students tended to accord these particular items a value about midway in their overall pattern of agreement-disagreement.

These data mean, to this author at least, that for the eight items listed in Table I, this particular group of prospective teachers rather accurately estimated adolescent response. They have some idea as to how high school students think and feel, at least as regards the ideas inherent in this particular set of items.

Concerning the differences apparent in Table II, there are probably several interpretations possible. Some which this author feels may be important are discussed below. Many of these have been projected to their logical extreme, and as such, may appear to go beyond the data. The author feels that these projections, if not warranted, are at least permissable on the grounds that they may serve to encourage further exploration into the area.

First of all, do the differences which seem to be apparent really exist? Did the college students who plan to become teachers really miss their estimate of high school students as much as seems to be the case from a study of the data? In other words, are these differences really differences, or are there some factors which have operated to produce what appear to be several discrepancies whereas actually the differences are not real; they do not exist?

Second, if these differences actually do exist, why do they? That is, what accounts for the fact that this group of prospective teachers has such a faulty perception of high school students?

Regarding the first question, "Do the differences really exist?" there are at least two points deserving further attention. It may be that the high school students who participated in this study responded to the various items as they thought they were supposed to respond. This is a point, very obviously, which is always quite difficult to ascertain with accuracy in any study of this kind. It may be, also, that these high school students really know or feel one way (i.e., as they responded) but actually behave in an entirely different manner. Of course, if this is so, the college students' estimate might actually more closely approximate adolescents' real behavior, which is observable, but then the question immediately arises: "Which is the important thing, how adolescents feel or how

TABLE I

PROSPECTIVE TEACHERS' CORRECT ESTIMATES OF HOW HIGH SCHOOL STUDENTS WOULD RANK F SCALE ITEMS

Item Number	Item	Actual High School Ranking	College Estimate of H. S. Ranking
2	No weakness or difficulty can hold us back if we have enough will power.	3	3
3	Science has its place, but there are many things that can never possibly be understood by the human mind.	6	7
18	Nowadays more and more people are prying into matters that should remain personal and private.	7	5
4	Human nature being what it is, there will always be war and conflict.	13	11
9	Nowadays when so many different kinds of people move around and mix together so much, a person has to protect himself especially carefully against catching an infection or disease from them.	16	18
27	Familiarity breeds contempt.	20	19
11	Some people are born with an urge to jump from high places.	21	20
26	Nobody ever learned anything really important except through suffering.	27	25

TABLE II

PROSPECTIVE TEACHERS' INCORRECT ESTIMATES OF HOW HIGH SCHOOL STUDENTS WOULD RANK F SCALE ITEMS

Item Number	Item	Actual High School Ranking	College Estimate of H. S. Ranking
1	Obedience and respect for authority are the most important virtues children should learn.	1	27
12	Young people sometimes get rebellions ideas, but as they grow up they ought to get over them and settle down.	2	24
17	Someday it will probably be shown that astrology can explain a lot of things.	4	23
16	There is hardly anything lower than a person who does not feel a great love, gratitude, and respect for his parents.	5	16
8	What youth needs most is strict discipline, rugged determination, and the will to work and fight for family and country.	8	26
22	If people would talk less and work more, everybody would be better off.	12	22
7	A person who has bad manners, habits, and breeding can hardly expect to get along with decent people.	17	7
6	When a person has a problem or a worry, it is best for him not to think about it, but to keep busy with more cheerful things.	19	9
20	Most of our social problems would be solved if we could somehow get rid of the immoral, crooked, and feeble-minded people.	23	14
15	People can be divided into two distinct classes, the weak and the strong.	25	12
10	An insult to our honor should always be punished.	22	1
24	The business man and the manufacturer are much more important to society than the artist and the professor.	26	4

behave? Certainly there are arguments for both sides of any such question.

Another criticism of the data in this study may be that the samples are not "representative," or at least that the group of college students from Pennsylvania need not necessarily be expected to know how high school students from Michigan and Alabama would feel regarding these items of the F scale. Their estimates might actually be correct for a group of high school students from their own geographical or cultural vicinity, for example. This is, of course, an important point, but one which can only be resolved with further study. One might infer that since the Michigan and Alabama samples were actually quite similar in their relative agreement with the various items of the measure, Pennsylvania high school students would probably closely approximate this pattern of agreement-disagreement too, but this could never be more than speculation, however carefully concluded.

However, disregarding the limitations for the time being, if these differences really do exist, a very real question is "Why?" Why do adolescents feel one way, and why do college students think they feel another way? Are the prospective teachers learning the wrong things in their college courses and/or from the media of mass communication, radio, press, television, .etc.? Further, are the high school students actually as "good" as an inspection of the data seem to indicate? Do prospective teachers really think that high school students are as "bad" as they seem to think if one examines their estimates of high school students' responses? Then, if they actually do think that high school students are that "bad," why are they going into teaching? What are their motives regarding entering a profession which will entail working with whole groups which they evidently feel are essentially "bad?"

Considering all of these factors, the differences which seem to be apparent in Table II will be discussed in light of the opportunities they suggest for further research. In other words, there seem to be some factor or factors involved here which result in certain differences which have manifested themselves in the data contained in Table II. However, since the data and the devices employed in this study have certain definite limitations, the author feels that each of these various areas deserves further extensive investigation.

Suggested Research

Do adolescents really feel that "obedience and respect for authority are the most important virtues children should learn?" Does this mean that they want, like, and even expect direction, control, and limits? Further, do prospective teachers feel that high school students are really "authority rebels," that they dislike control, and that they do not want to show respect for authorities of any kind, as the data in this particular study seem to imply?

Are adolescents really concerned about people of all kinds? Do they really feel that people are important, as these data seem to indicate? And do college students who intend to teach actually think that high school students will want to "get rid of" certain groups, even those persons who might be conveniently classed as "undesirables?"

Finally, do these data mean that high school boys and girls are not aggressive, do not follow an "eye for an eye" kind of philosophy, and will not demand redemption for wrongs committed against them? It this the counterpart of the "love and respect" and "all people count" ideas which seem to emerge from their responses to the ideas inherent in the several items as they agreed or disagreed with them? Also, do prospective teachers really think that high school students function on a strict "one for one" basis, punishing insults and rejecting their parents, as might be inferred from their estimates regarding adolescent responses to the various items?

Conclusions

Obviously there are many conclusions one might draw from these kinds of data. Four things seem especially important to this writer. First the general interpretation of the data seem to indicate that high school students' values, as reflected indirectly in a ranking of the various items of the F scale according to their agreement, are positive. They seem to think, in other words, that "man is basically good."

Second, it would seem that by and large this group of prospective teachers' estimates of high school students is negative. They seem to think, in other words, that high school students think "man is basically bad."

Third, although there is a rather sizeable area of values which the prospective teachers correctly estimated, in general their incorrect estimates seem more significant and more profound by virtue of the very items involved in error. Said another way, this group of young people, who not only hope but are seriously planning and working toward a teaching career, have an essentially inaccurate perception of what adolescents are actually like.

Finally, two questions beg of further study. Are the differences which are apparent here really differences? Second, if they do exist, why do they?

Summary

Two groups of young people, 104 high school students from Michigan and Alabama and 69 college students enrolled in a Teachers College in a Pennsylvania university were measured with the F scale. The way the high school students agreed and disagreed with the various items of the measure was compared to the way the college students thought high school students would agree and disagree. The prospective teachers correctly estimated adolescent responses for eight of the 27 items, but incorrectly estimated their responses for 12 items. Several problems were posed for further investigation.

REFERENCES

1. Scodel, A. and Mussen, P., "Social Perceptions of Authoritarians and Non-Authoritarians," Journal of Abnormal and Social Psychology, XLVIII (April, 1953), 181-184.
2. Cohn, T. S., "Is the F Scale Indirect? " Journal of Abnormal and Social Psychology, XLVII (July, 1952), 732.
3. Frymier, Jack R., "Analysis of Adolescents' Responses to the F Scale," Journal of Expermental Education, LX (Sept., 1960), pp. 73-80.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 2, December 1960)

PERSONALITY TRAITS ASSOCIATED WITH MATHEMATICAL ABILITY: A PLEA FOR RESEARCH

ROBERT H. RIFFENBURGH
University of Hawaii

IN RECENT years increasing attention has been given to inadequacies in training of mathematicians in our society. With ludicrous frequency one hears it said that the student in question does not have a "mathematical mind." What is this mathematical mind? Does it exist? If so, early recognition would permit proper attention to promising individuals and could reduce the cost and frustration of attempting to train the untrainable. If the mathematical mind does not exist, revision of our attitudes and instruction is just as urgent.

A review of the pertinent literature reveals that existing studies give disappointingly narrow and mostly negative answers to the above questions.

In this paper, a mathematical mind is assumed to be an innate phenomenon, unrelated to basic or general intelligence,* existing prior to mathematical training, which enables the individual possessing it to learn and perform mathematics with facility. The name could well be "mathematical personality."

Extensive consideration of the subject by various mathematicians, psychologists, and educators suggested that the following five factors of the ability to perform mathematics be postulated:

1. general intelligence
2. past training
3. general maturity
4. attitude toward and interest in mathematics
5. a mathematical personality

Prior studies add support to some of these postulates. Cattell (3) performed a factor analysis on results of Army Alpha Intelligence Tests and the verbal and mathematical sections of the Graduate Record Examination. He found the factors (a) general ability (similar to 1), (b) character integration (may be related to 3), and (c) trained mind or length of education (similar to 2); he also found some evidence of "cleverness, analytical mind," which might be related to any of 1, 2, 3, or 5. Barakat (1) gave a set of mathematics tests to youths and performed a factor analysis on the results. He found the factors (a) general intelligence (same as 1); (b) group factor for mathematical ability (which from his description depends on past training, 2); (c) verbal; (d) visuo-spatial factor. (c) and (d) might relate to any of 1, 2, 3, or 5. The reader is reminded that the naming of factors obtained from factor analysis is somewhat arbitrary.

In order to investigate the existence or nonexistence of factor 5, it is desirable to eliminate by some means or other the first four factors and examine what remains. Five general approaches have occurred to the author.

Longitudinal Studies—The longitudinal study promises the most information of all studies, but presents by far the most difficult problems. Many children would be followed from an early age to adulthood with intermittent testing throughout. As adults, those doing well in mathematics could be compared with those doing poorly on the basis of heredity, continuous environment, and psychologic measures of all sorts.

Pros: Information would be more complete, more certain, and better controlled than in other approaches. Given large enough samples, intelligence could be equated, subjects could be tested prior to any training, training could be controlled, and comparisons could be made in maturity. Thus factors 1, 2, 3, and to some extent 4, could be eliminated from consideration.

Cons: The obvious difficulties are the length of time needed, the enormous cost involved, and the problem of obtaining and retaining subjects. Also of importance is the possibility that the controls themselves—attention to personality and performance, frequent testing, etc., —may exert an influence upon some members of the group.

There have been no studies to date using this approach.

---

*Witness the many quite intelligent social scientists who claim they have "no mathematical mind."

Child Studies—The child study approach would involve equating for intelligence and testing and comparing a large sample of children on the basis of personality tests and some mathematical device.

Pros: There will have been no mathematical training and the childrens' attitudes will not have been formed as yet; thus factors 2 and 4 are controlled effectively.

Cons: If maturity, factor 3, exists as a necessary condition for mathematical ability, and this has appeared to be the case in many instances, this method is useless. Furthermore, obtaining or devising a measure for children which will be associated with adult performance in mathematics is requisite and a very difficult task. A language factor is present and communication is difficult. Personality and intelligence measures for young children are not remarkably reliable.

There have been no studies to date using this approach.

Adult Studies—An adult study would involve (a) testing a sample of adults equated for intelligence, attitude, and past training, (b) assuming the maturity condition present, and (c) comparing results for those competent mathematically with those not competent mathematically.

Pros: Adults are easier to measure than children; maturity is much more likely to be present than in younger groups.

Cons: The maturity condition must still be assumed to be satisfied, or some measure of maturity must be made and only individuals satisfying the maturity condition be chosen; if this maturity involves "mathematical maturity," such a measure does not now exist. Equating groups for intelligence and attitude may be done to some extent through existing psychologic measures, but these measures have limitations; equating for past training would be extremely difficult. Furthermore, even if the mathematically able were found to have traits different from the mathematically unable, there arises the question: are the individuals able because of these traits or have they acquired some of these traits due to their abilities and professions?

There have been no studies to date using this approach.

"Eminent Men" Studies—Here the personalities of "eminent men" in the field of mathematics would be investigated in an effort to determine characteristics unique to them; this may involve comparison with "normal" men or with eminent men in other fields.

Pros: If personality traits peculiar to mathematical competence exist, the subjects from this group may be safely assumed to have them. Thus absence of differential results would be much stronger evidence of the non-existence of such personality traits than would similar absence in other types of subjects.

Cons: Two questions are posed. Would the eminent mathematicians have been as eminent if they had gone into another field? Do the individuals follow their professions because of their personalities, or have they acquired some of these traits due to their professions? Also, if comparison is made with "normal," factors 1, 2, and 4 may be expected to be present. If comparison is made with groups eminent in other fields, individuals in the other fields may be or may be potentially mathematically competent as well.

A few studies related to this approach have been made by Roe (5, 6, 7), but mathematics was not the primary interest and no results relating to mathematics were given. Roe compared groups of scientists, viz., physical, biologic, psychologic, and anthropologic scientists, on the basis of the Rorschach Test. Some results were found and were expressed as differences in frequency of various types of responses. The interpretation was generally left to the reader, apparently for the reason that interpretation is not unique and will vary from one psychologist to another. Roe also investigated a group of the very "eminent" in several fields of science and gave "case histories." The results may provide ideas as to traits to investigate, but little more.

Youth Studies—A youth study would be essentially a compromise between the child and adult studies, possessing some of the advantages and disadvantages of each. It would involve sampling from youths of high-school or college age. Firmly formed attitudes and effects of past training are less than in adults but more than in children; maturity would be more complete than in children but less than in adults. The intelligence factor could be largely eliminated by comparing, say, groups of college students in the same curriculum where one group was composed of A-average students receiving A in mathematics courses and the other of A-average students receiving F in mathematics courses.

Pros: The subjects would be readily available, since most researchers are affiliated with universities. Also the testing problem could be simplified by the universities' testing services.

Cons: Some of the disadvantages of both the child and the adult studies would be present.

Only one study using this approach has been done; the results were negative. The procedure was to sample college freshmen, comparing by a personality evaluation technique a mathematically capable group with a mathematically inept group. It was assumed that the maturity level was nearly adult and that the subjects were not trained mathematically to any degree which could affect their personalities. In order to reduce the effects of the past training and attitude factors, the sample was chosen from among engineering freshmen at one state school. In order to attempt to equate for intelligence, the groups were chosen from students in the upper half of their class; one group had received A in college algebra, the other F.

The personality test posed a problem. It was necessary to eliminate all factors but 5 from the measure. (Any attitude resulting from natural proclivity could be thought of as 5 rather than 4 and is desired anyway.) Of personality measures available to the author, all types but projective techniques may be affected by intelligence or attitude. Of possible projective techniques, it was felt necessary to choose one producing quantifiable results in order to obtain reproducible results susceptible to statistical analysis. Burgess (2) has indicated that quantification and statistical testing of the Rorschach is inadequate for differentiating academic deviates, and that the Rosenzweig Picture Frustration Study and the Minnesota Multi-Phasic Inventory group form are not at all suitable. She recommends the Thematic Apperception Test, but that technique presents serious problems in quantification. Furthermore, all the mentioned projective techniques lack unique interpretations; different psychologists often find different results. A number of lesser-known and foreign projective techniques were investigated and it was decided that the little-known Polychrome Index (4, 8, 9, 10) would most nearly satisfy the requirements. This index is brief and completely quantifiable. It has been shown (8, 9) to be unrelated to intelligence; it appears safe to assume that it is unrelated to past training or attitude toward mathematics. This index may not be, however, unrelated to the maturity factor, but in fact has been shown (4) to relate closely to certain aspects of "social maturity."

For the Polychrome Index, nine plates containing various color combinations are presented and the subject is asked to choose the three he likes most (ordered), which are then weighted 3, 2, 1, and the three he likes least (ordered), which are weighted -3, -2, -1; the three not chosen are weighted 0.

The Polychrome Index was administered in the standard fashion to 111 college engineering freshmen at the Virginia Polytechnic Institute enrolled in college algebra. At the end of the term the scores for 22 students receiving the grade "A" and the 21 receiving "F" were isolated and compared by t tests.

The table shown on this page gives the mean scores for A students and F students for each plate on the Polychrome Index; it gives the t's computed for each plate comparing the means of the A and F students. It is seen that no t is significant.

Plate	Mean Score for "A" Students	Mean Score for "F" Students	t*
I	1.86	1.24	1.42
II	.23	.10	.02
III	.09	-.14	.10
IV	1.09	1.14	.12
V	-.41	-.90	1.78
VI	-.05	-.29	.12
VII	-.45	0.00	.85
VIII	-.14	-.41	.45
IX	-1.50	-.81	1.80

*No t's significant.

There are three possible interpretations of the negative results given in the table:

a) There is no "mathematical personality."
b) There are several different "mathematical personalities."
c) There is a "mathematical personality," but the investigation was not adequate to find it.

We may conclude that here is presented some small but certainly not conclusive evidence that there is no "mathematical personality." If repeated studies should give similar findings, we may develop some confidence that there is no mathematical personality and consequently no "mathematical mind."

Summary

The question, "Is mathematical ability innate?" is becoming of urgent importance in our contemporary culture. The training of competent mathematicians would be markedly facilitated by finding the answer to this question. No research of import or value has been done that provides information about this problem.

This paper presents a plea for research on this subject. It presents five approaches to such research, discussing the advantages and disadvantages of each and mentioning the few studies that have bearing upon the field. One such study, done by the author, is given in the preceding section of of this paper.

REFERENCES

1. Barakat, M. K. "A Factorial Study of Mathematical Abilities," *British Journal of Statistical Psychology, IV (1951), pp. 137-56.*
2. Burgess, E. *Personality Factors in Over- and Under-Achievers in Engineering, Penn State University Abstracts of Dissertations,* 1953-1955, XVI (1955), pp. 536-43.
3. Cattell, R. B. "Personality Traits Associated with Abilities: with Verbal and Math Abilities," *Journal of Educational Psychology, XXXVI (1945), pp.* 475-86.

4. Riffenburgh, G. H. "Response to Color Combinations as Indices of Personality Traits," Journal of General Psychology, LXI (1959), pp. 317-22.
5. Roe, A. "A Psychologist Examines 64 Eminent Scientists," Scientific American, CLXXXVII (1952), pp. 21-5.
6. Roe, A. "Analysis of Group Rorschachs of Psychologists and Anthropologists," Journal of Projective Techniques, XVI (1952).
7. Roe, A. "A Psychological Study of Eminent Psychologists and Anthropologists and a Comparison with Biological and Physical Scientists," Psychological Monographs, LXVII (1953).
8. Woods, W. A. "Personality Through Color," Mental Health in Virginia (Summer 1954), pp. 51-4.
9. Woods, W. A. "Some Determinants of Attitudes Toward Colors in Combinations," Perception and Motor Skills, VI (1956), pp. 187-93.
10. Woods, W. A. A Polychrome Index of Social Maturity, paper read at the Chicago meeting of the American Psychological Association, September 4, 1956.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 2, December 1960)

FACTORIAL CONTENT OF THE IOWA TESTS OF EDUCATIONAL DEVELOPMENT AND OTHER TESTS

RUSSELL N. CASSELL
Lompoc, California
EDWARD J. STANCIK
Phoenix, Arizona

THIS STUDY was concerned with isolating by centroid factor analysis the independent factors present in the Iowa Tests of Educational Development (ITED), and certain other tests and subtests; and to identify and describe such factors on the basis of face and construct validity by rotating them orthogonally to simple structure. A total of 15 variables in the form of tests and subtests were included in the factor analysis, seven of which were not a portion of the ITED battery.

Tests Utilized--Achievement Area

All of the tests and subtests included in this study were a portion of the regular test battery administered for guidance purposes to all entering 9th grade (freshmen) students in Phoenix high schools.

Iowa Tests of Educational Development X-3S ITED). These tests were administered during freshmen week (immediately prior to the opening of school in the fall) by guidance counselors in the various guidance departments of the respective high schools. Eight of the tests from this battery were included in the factor analysis: 1) Basic Social Concepts, 2) Natural Science Background, 3) Correctness of Expression, 4) Quantitative Thinking, 5) Interpretation of Social Studies, 6) Interpretation of Natural Science, 7) Interpretation of Literature, and 8) Vocabulary. The standard scores were used, and they were compiled from national first semester freshmen norms for the 1957-58 school year.

Cooperative English-Reading Comprehension Cl - (Y) Lower Level. This test with its several parts (Vocabulary, Speed, and Level) was administered in the same manner as the ITED. Scaled standard scores were utilized based on national norms provided in the test manual.

Tests Utilized--Multiple Aptitude Area

Differential Aptitude Test Battery (DAT). This battery was administered by staff members and teachers of the 12 elementary school districts that channel students into the Phoenix Union High Schools and Phoenix College System. Generally, the tests were administered during the first half of the eighth grade school year (1956-57). Four of the tests from the DAT were included in the factor analysis: 1) Verbal Reasoning, 2) Numerical Ability, 3) Language Usage-Spelling, and 4) Language Usage-Sentences. Raw scores were utilized.

Tests Utilized--Intelligence Area

California Test of Mental Maturity, Advanced, '57 S-Form (CTMM). This test was administered during freshmen week to all entering 9th grade students by the guidance counselors of the respective high schools. Both parts of the CTMM were included in the factor analysis, and in terms of a raw score for each: 1) Language Scale and 2) Non-language Scale.

Group Utilized

There were 4,776 9th grade entering students for the 1957-58 school year, and the 124 students selected for this study were screened on the basis of a two-fold process: 1) random selection, and 2) elimination of non- or poor readers.

Every fifth name from an alphabetical listing of entering 9th grade students in each of the seven high schools was included for the initial portion of the study (20 percent sample of total group). This group was then further reduced by a second selection of every fifth name (20 percent sample of the initial 20 percent sample). This group was then further screened to eliminate those individuals with a Cooperative Reading stanine score of three or below. This elimination of poor readers was accomplished to insure that all individuals included in the factor analysis study (the 124 individuals used for study) were capable of reading

TABLE I

INTER-CORRELATION MATRIX OF THE FIFTEEN VARIABLES TOGETHER WITH THE MEANS AND STANDARD DEVIATIONS FOR THE 124 INDIVIDUALS OF STUDY (N-124)

Tests and Variables	Inter-correlations Pearson r's															Mean	SD
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}		
1. California Test of Mental Maturity, Advanced '57 S-Form																	
X_1 Language Scale Score		426	555	428	337	459	474	535	499	444	489	534	228	506	523	28.57	7.08
X_2 Non-language Scale Score			439	429	176	433	309	385	315	257	446	263	170	278	357	36.98	4.45
2. Differential Aptitude Test Battery																	
X_3 Verbal Reasoning				508	379	546	672	585	581	520	587	735	617	517	642	16.93	7.40
X_4 Numerical Ability					398	525	436	409	377	515	502	466	399	373	484	13.73	6.65
X_5 Language Usage-Spelling						469	466	209	387	675	291	415	332	367	438	29.59	23.24
X_6 Language Usage-Sentences							506	435	368	635	438	517	443	630	559	24.84	12.71
3. Cooperative English-Reading Comprehension, Cl-(Y) Lower Level																	
X_7 Total Score								582	552	564	604	748	639	699	721	44.34	7.45
4. Iowa Tests of Educational Development, Y-35																	
X_8 Basic Social Studies									691	423	618	711	603	581	652	13.00	6.10
X_9 Basic Natural Science										240	542	683	498	551	632	13.30	5.30
X_{10} Correctness of Expression											479	499	442	552	561	12.30	4.94
X_{11} Quantitative												582	509	381	482	12.60	5.57
X_{12} Interpretation of Social Studies													744	694	733	12.80	4.93
X_{13} Interpretation of Natural Science														542	598	12.18	5.91
X_{14} Interpretation of Litarature															718	11.63	5.33
X_{15} Vocabulary																12.81	5.29

r's of 238 or better significant at 1% level.

r's of 174 or better significant at 5% level.

TABLE II

CENTROID AND ORTHOGONAL ROTATED FACTOR LOADINGS TOGETHER WITH COMMUNALITIES (N-124)

	Centroid Factor Loadings						Rotated Factor Loadings						Communalities	
	I	II	III	IV	V	VI	I	II	III	IV	V	VI	h_c^2	h_r^2
1. California Test of Mental Maturity Advanced '57 S-Form														
X_1 Language Score	654	-155	236	-213	-114	201	095	491	287	-195	444	194	606	605
X_2 Non-language Score	480	-257	352	064	-259	-073	-096	554	226	031	336	-136	497	500
2. Differential Aptitude Test Battery														
X_3 Verbal Reasoning	806	041	110	121	-040	-063	054	289	229	159	719	049	684	684
X_4 Numerical Ability	634	-284	059	150	-038	-112	156	384	373	226	374	-144	523	523
X_5 Language Usage-Spelling	572	-306	-315	-182	305	-194	569	-049	513	034	234	-184	684	679
X_6 Language Usage-Sentences	711	-244	-179	089	-257	-045	391	509	217	249	376	-117	673	676
3. Cooperative English-Reading Comprehension C1-(Y) Lower Level														
X_7 Total Score	813	163	-142	-045	049	100	278	169	159	084	712	276	722	721
4. Iowa Tests of Educational Development Y-35														
X_8 Basic Social Studies	761	247	271	079	069	181	-129	171	163	032	761	332	757	763
X_9 Basic Natural Science	712	280	265	-262	136	-181	-011	004	162	-264	823	009	775	773
X_{10} Correctness of Expression	700	-346	-387	081	090	135	574	279	427	293	253	234	792	794
X_{11} Quantitative	717	- 09	246	131	277	099	-050	190	554	135	540	184	690	689
X_{12} Interpretation of Social Studies	849	331	-031	044	049	044	119	079	074	089	857	259	838	835
X_{13} Interpretation of Natural Science	702	348	-160	358	160	-118	069	-149	064	414	764	124	807	802
X_{14} Interpretation of Literature	758	178	-252	-160	-219	166	402	286	-230	-079	624	284	771	773
X_{15} Vocabulary	826	181	-096	-129	-122	-033	274	249	044	-093	769	133	757	757

Note: Decimal points omitted-loadings of 300 or more considered statistically significant at 1% level.

the tests included.

A comparative examination of mean test scores for intelligence and reading indicated that the average for the students in the final experimental group was significantly higher than the average for the total freshmen group.

Findings

Six separate and independent centroid factors were extracted from the 15 different tests and subtests contained in this study. Each of these six factors has been identified on the basis of the significant factor loadings (values of .300 or better considered to be statistically significant at the one percent level).

Factor I--Correctness of Expression-English. This factor was relatively pure in terms of the significant loadings present, and with the Correctness of Expression Test of the ITED being the most representative and characteristic variable in the study. Other significant factor loadings were: Language Usage-Spelling (.569), Interpretation of Literature (.402), and Language Usage-Sentences (.391), see Table II. Each of these tests have face validity that agrees with the defined factor, and they all have high agreement with each other.

Factor II--Intelligence-Language and Non-Language. This factor appeared to be relatively pure also, in that all four variables (tests) having significant loadings were from either the aptitude or intelligence area. The fact that both parts from the only intelligence test included in the analysis (CTMM) had significant loadings adds credence to the CTMM for assessing this factor. Other tests which had significant loadings on this factor were as follows: Language Usage-Sentences (.509), and Numerical Ability (.384).

Factor III--Quantitative. This factor did not appear to be factorially pure in terms of the face validity for the tests having significant loadings. The fact that both the Quantitative Thinking test of the ITED and the Numerical Ability test of the DAT obtained significant loadings adds credence to referenced identity (these loadings were .554 and .373, respectively). The first of these two variables is recommended for describing the factor, i. e., Quantitative Thinking.

It is of interest to note that the other two tests having significant loadings on this factor were: Language Usage-Spelling (.513), and Correctness of Expression (.427). The implication is that effective quantitative performance entails some aspect of language competency, and this may be true because written problems are included in the tests.

Factor IV--Natural Science. This factor was factorially pure in terms of significant factor loadings, and the Interpretation of Natural Science test from the ITED [illegible] the only variable having significant weight (.414).

Factor V--Reading Competency. In spite of extensive rotations, simple structure was not obtained for this factor, and it could not be considered to be factorially pure in any sense of the statistical concept. In terms of face validity for reading competency, the Interpretation of Social Studies (.857) from the ITED, and Cooperative English-Reading Comprehension (.712) are recommended for the identification of this factor. The effective reading function of materials on the high school level obviously involves many of the variables included in the various tests of this study.

Factor VI--Basic Social Studies. This factor was factorially pure in terms of significant factor loadings, and the Basic Social Studies test of the ITED was the only variable that had significant loadings (.332).

Summary

This study was concerned with isolating by centroid factor analysis the independent factors present in the Iowa Tests of Educational Development (ITED) and certain other tests and subtests, all of which form a portion of the 9th grade guidance test battery of the Phoenix Union High Schools and Phoenix College System. Six centroid factor loadings were extracted, and five of these were rotated to simple structure orthogonally. These six factors were identified as follows, and the indicated tests were recommended for their assessment in the Phoenix Union High Schools: I-Correctness of Expression-English (Correctness of Expression from ITED); II-Intelligence-Language and Non-language (California Test of Mental Maturity); III-Quantitative Thinking (Quanitative Thinking from ITED); IV-Natural Science (Interpretation of Natural Science from ITED); V-Reading Competency (Interpretation of Social Studies from ITED); and VI-Basic Social Studies (Basic Social Studies from ITED).

REFERENCES

1. Cronbach, L. J., and Meehl, R. E., "Construct Validity in Psychological Tests," Psychological Bulletin, July, 1955, Vol. 52, No. 4.
2. Thurstone, L. L., Multiple Factor Analysis, Chicago: University of Chicago Press, 1947.
3. Zimmerman, W. S., "A Simple Graphic Method for Orthogonal Rotation of Axes," Psychometrika, XI (1946).

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 2, December 1960)

THE HOUSE-DRAWING TEST AS A PREDICTOR OF FIRST-GRADE SCHOOL ACHIEVEMENT

HARRY S. BECK
Iowa State Teachers College, Cedar Falls, Iowa[1]*
ROBERT C. BECK
Wake Forest College[2]

PREVIOUS STUDIES of young children's house drawings have suggested the possibility of using these as a readiness test for primary school. Beck (1955) and Markham (1954), for example, found differences in the drawings of five- and six-year-olds which seemed to differentiate between dull and superior children. The present study was undertaken to determine if any of these characteristics, easily and quickly obtainable from entering first-graders, are related to first-grade achievement.

Method

Subjects. The subjects consisted of the entire first-grade populations of West Frankfort and Benton, Illinois, two communities located about seven miles apart. There were twelve first-grade classes, but the data from one of these was unduly incomplete and therefore was not used in the present analysis. The number of subjects reported herein is 214. All Ss were white, there being no negroes in either community.

Procedure. The teachers were well briefed on the project and all tests were administered by them. In September, the California Test of Mental Maturity (CTMM) was given and the house drawings were obtained. In late December, the Haggerty-Olsen-Wichman Behavior rating scales A and B (HOW) were completed by the teachers, and in May of the following spring, the American School Achievement Test-Primary I Battery (ASA) was given.

The Drawings. Each teacher was given individual written and verbal instructions for the administration and scoring of the drawings. To increase scoring reliability, one of the authors (HSB) met with the teachers to explain the task to them and to clarify the meaning of the various characteristics upon which the drawings were to be checked.

At the time of administration, the teacher gave each child in her class a sheet of plain white paper 7 x 8.5 in. (a half sheet of legal-size paper), number two pencils and erasers. She was to see that there were no pictures of houses visible which the children could copy, and that each child had his paper placed with the long edge horizontal to his line of vision. She then gave the following directions: "You all have paper and pencils and we are going to draw a house. Draw the very best house you can. I want to see if you can draw better houses than the other children did."

The drawings were scored simply on the presence or absence of fifteen characteristics which previous research and experience indicated might be related to developmental level. Seven of these were "negative" (indicators of "immaturity") and eight were "positive" (indicators of "greater maturity"), relative to the performance of the average six-year-old. These indicators were as follows:

Negative Characteristics

1. Omission of more than one essential detail (roof, wall, door, window, or chimney).
2. Poor proportion (e. g. door wider than high, house that looks like a tower.)
3. Poor motor coordination (e. g. lines forming walls are wavy or jagged, corner lines overshoot each other).
4. An unrecognizable drawing.
5. A recognizable but bizarre-appearing drawing.
6. Transparencies (showing details behind solid walls--the "glass-house" effect.
7. Poor organization (e. g. windows not in line horizontally or vertically, or random place-

---

*Footnotes will be found at the end of the article.

ment of a door in the center of a wall instead of at the base.

Positive Characteristics

1. Two walls instead of one.
2. Three or more windows.
3. Correct alignment of windows and doors.
4. A window in the attic.
5. Two-dimensional roof.
6. Door knob.
7. Panes or curtains in the windows.
8. Door panels.

Each drawing was scored by the administering teacher and one of the authors (HSB) who had used such drawing tests extensively. In addition to the fifteen individual check items, two additional scores were obtained by summing the number of positive and negative checks.

Analysis of Data

All of the scores thus obtained, plus the chronological age of each subject, were put into a 43-variable correlation matrix with the 214 subjects from the eleven classes combined. This matrix was then factor analyzed by the centroid method, putting unities into the diagonal spaces. There were four rotations to simple structure using the oblimax solution (Pinzka and Saunders, 1954), rotating the first three, four, five, and six factors respectively. All computations were done on the University of Illinois high speed digital computor, Illiac.

Results

As determined by simple analyses of variance the eleven subgroups were not significantly different on the objective measures, so all of the subgroups were combined into one large group. The means and standard deviations for the objective measures were then as follows: Age: 6.47 and .69; CTMM-total: 111.62 and 18.97; ASA-total: 7.06 (age equivalent) and .69; HOW-Scale A: 15.51 and 19.60 (highly skewed) and HOW-Scale B: 67.59 and 21.07. Since the drawings were rated by different teachers the reliability of the drawing scores from this source might be lowered in the combined matrix, but the resulting correlations suggest no particular gain or loss from this procedure.

Table I is the correlation matrix for the objective test scores and ratings, age, and the teachers' and psychologist's summed scores for positive and negative characteristics in the drawings. It should be pointed out that a high HOW score is "bad", i.e., indicative of a behavior problem. Almost all of the correlations in Table I are statistically significant, the magnitude of correlation necessary for significance being given in the footnote to the table.

The summed scores were almost invariably more highly correlated with the objective measure than were individual items, but a goodly number of individual house-drawing items actually were significantly correlated with intelligence, achievement and the HOW-scales. None were significantly correlated with age. While these correlations were in the proper direction, none of them had higher absolute value than about .30.

To determine if the drawing scores added to the correlation of .51 between the CTMM total and the ASA-total, four multiple correlations were computed, using each of the four summed scores as an additional predictor variable.[3] One of these multiple correlations was .52 and the other three were .51. Clearly, the drawing scores as used here do not contribute anything to the prediction of achievement which cannot be obtained by the CTMM alone.

Factor Analysis

Fifteen factors were extracted from the correlation matrix, accounting for 74 percent of the total variance. The first six factors accounted for 22.4, 7.8, 5.4, 4.6, 4.5, and 4.4 percent of the variance respectively, cumulating to a total of 49.1 percent. Each of the remaining factors accounted for only 2.3 to 3.3 percent of the variance.

For our purpose the six-factor rotation is the most useful since it gave one factor on which the CTMM scores loaded above .90 and another on which the ASA scores loaded at the same level. None of the four summed scores loaded higher than .18 on either of these factors, although one positive drawing characteristic (two walls instead of one) loaded .41 on the "achievement" factor. There was no large common factor among all the items, as we might guess from looking at the proportion of the variance accounted for by the individual factors. In brief, then, this analysis revealed only a conglomeration of factors which had only one or two or three drawing items loading on each, and with only a single item related to achievement.

Discussion

Objective Test Scores. In Table I, the intercorrelations of the objective tests tend to run high. The positive correlation between intelligence and achievement was predictable, but the substantial negative correlation between the CTMM scores and the HOW scores, and to a lesser extent between the ASA and HOW was less expected. It is likely that less intelligent students would be less adaptable in conforming to "acceptable social patterns" at this age, but the interpretation of the results may not be quite this simple. The teach-

TABLE I

CORRELATION MATRIX FOR OBJECTIVE TESTS AND SUMMED DRAWING SCORES

(Items 10-13 are teachers' and psychologist's summed scores for positive and negative drawing characteristics)

	1	2	3	4	5	6	7	8	8	10	11	12
1. CTMM (language)												
2. CTMM (non-language)	.65											
3. CTMM (total)	.89	.90										
4. Age	-.14	-.18	-.18									
5. ASA (reading)	.42	.43	.48	.25								
6. ASA (numbers)	.41	.41	.46	.35	.67							
7. ASA (total)	.44	.46	.51	.29	.87	.81						
8. HOW (Scale A)	-.42	-.39	-.46	.10	-.36	-.20	-.27					
9. HOW (Scale B)	-.46	-.46	-.53	.27	-.29	-.13	-.23	.68				
10. Teacher (neg.)	-.33	-.28	-.35	.00	-.22	-.27	-.24	.31	.27			
11. Teacher (pos.)	.24	.28	.31	.05	.15	.17	.14	-.27	-.26	-.59		
12. Psychologist (neg.)	-.39	-.37	-.42	.02	-.31	-.25	-.28	.31	.23	.54	-.47	
13. Psychologist (pos.)	.29	.26	.31	.03	.18	.20	.17	-.27	-.29	-.57	.82	-.53

Note: - for N = 214, a correlation of .14 or greater is significant at the .05 level and a correlation of .18 or greater is significant at the .01 level of confidence.

ers were familiar with the work of their students as well as probably knowing something of their C TMM scores. Consequently, their behavior ratings may have been flavored by this additional knowledge. The fact is, however, that the teachers did perceive the behavior of the students to be more troublesome in relation to intelligence.

Drawing Ratings--Rater Reliability. Since the teacher or psychologist rated a given drawing on both positive and negative characteristics, and these characteristics were expected to be more or less mutually exclusive in a given drawing, it is not surprising to find substantial negative correlations between the teacher's two sets of summed scores and the psychologist's two sets of summed scores (Table I). These are not overly high, however (-.59 and -.53), indicating a considerable number of both kinds of indicators in many of the drawings. A more important comparison is that for the ratings between the two sources. The psychologist's and teachers' ratings were correlated +.54 for summed negative characteristics and +.82 for summed positive scores. The latter correlation suggests a high level of rater reliability, particularly since there were eleven different teachers included in this correlation. Since the psychologist's correlations with the objective measures were almost exactly the same as the teacher's correlations, any differential scoring skills were not reflected; while the two rating sources differed, they were equally "good".

Drawings and Test Scores. The summed scores were significantly correlated in the proper directions with intelligence and achievement. Unfortunately, the correlations were not high enough to be of much use by themselves and the multiple correlations dispelled the notion that the drawings might add something significant to the correlations between intelligence and achievement. While some other carefully standardized drawing test, such as the Goodenough, might be useful for achievement prediction, we must conclude that the present test is not. Since this test had only a conglomeration of factors, with only one of the items loading on an "achievement" factor, neither does its potential for this purpose look good. One might in fact question whether such a "test" has any meaning at all.

Summary and Conclusions

The present investigation was concerned with the potential use of a house-drawing test for screening incoming first-graders. The problem was to determine if this device would have enough predictive validity to be useful in selecting out children who might have difficulty with school work, i.e., be low achievers. Over the school year an intelligence test, a behavior rating scale, a house-drawing test, and an achievement test were administered to 214 first-graders. The house drawings were scored on the presence or absence of seven "negative" characteristics (indices of "immaturity") and eight "positive" characteristics (indices of "greater maturity") by a child's teacher and by a school psychologist. A total of 43 variables (9 objective scores and 34 house-drawing scores) were put into a correlation matrix which was factor analyzed and rotated to a simple structure. The results were as follows:

1. There was a positive relationship between achievement and intelligence, and behavior was perceived by the teachers as being troublesome in relation to intelligence.
2. Many of the house-drawing characteristics were significantly correlated with intelligence and achievement, but these correlations were of low order, i. e., ranging in absolute value from .14 to .35.
3. Multiple correlations between the summed house-drawing scores, intelligence and achievement were no higher than those between intelligence and achievement alone.
4. The factor analysis revealed a factor for "intelligence" and one for "achievement", but only a single drawing characteristic loaded on either of these and there was no important common factor among the drawing scores themselves.
5. It is concluded that the house-drawing test, as used here, is not suitable for a screening device for selecting out entering first-graders who might have school achievement problems.

FOOTNOTES

1. Director, Educational Clinic.
2. The authors wish to acknowledge the assistance of Mr. Kern Dickman and members of the University of Illinois digital computing laboratory in carrying out the data analysis. Appreciation is also expressed to the superintendents of schools in the towns where data was collected, L. Goebel Patton and H. E. Boyd, and to their first-grade teachers.
3. At the suggestion of Dr. Jum C. Nunnally.

REFERENCES

Beck, Harry S. "A Study of the Applicability of the H-T-P to Children With Respect to the Drawn House," *Journal of Clinical Psychology*, 1955, 11, 60-63.

Markham, Sylvia. "An Item Analysis of Children's Drawings of a House," *Journal of Clinical Psychology*, 1954, 10, 185-187.

Pinzka, C. and Saunders, D. R. *Analytic Rotation to Simple Structure II: Extension to an Oblique Solution.* Research Bulletin RB-54-31, Educational Testing Service. Dated August 16, 1954.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 2, December 1960)

A DEVICE FOR QUICKLY ADDING WEIGHTED SCORES

JAMES W. CREASER
University of Illinois*

TO DETERMINE the predicted grades or other scores from a regression equation of the form

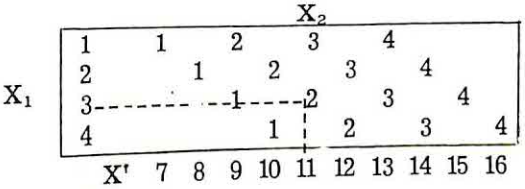
$$X' = AX_1 + BX_2 + C$$

a table or graph can be made up to eliminate the repetition of many calculations. By entering such a chart with the variable scores, the predicted score is found directly.

For example, if in the above equation

$$A = 1$$
$$B = 2$$
$$C = 4$$

with values of X ranging from 1 to 4, a chart could be made up as follows:



The broken lines indicate the path by which to find the predicted score for an individual with X_1 and X_2 scores of 3 and 2 respectively.

In making up such a chart, there is a limitation as to the number of variables, the range of scores and the complexity of the weights. The limitation is not in theory but in space. A chart too large would lose its efficiency.

Negative weights can be represented by putting the numbers for that variable into the chart in reverse order. The distance between numbers is determined by the magnitude of the weights.

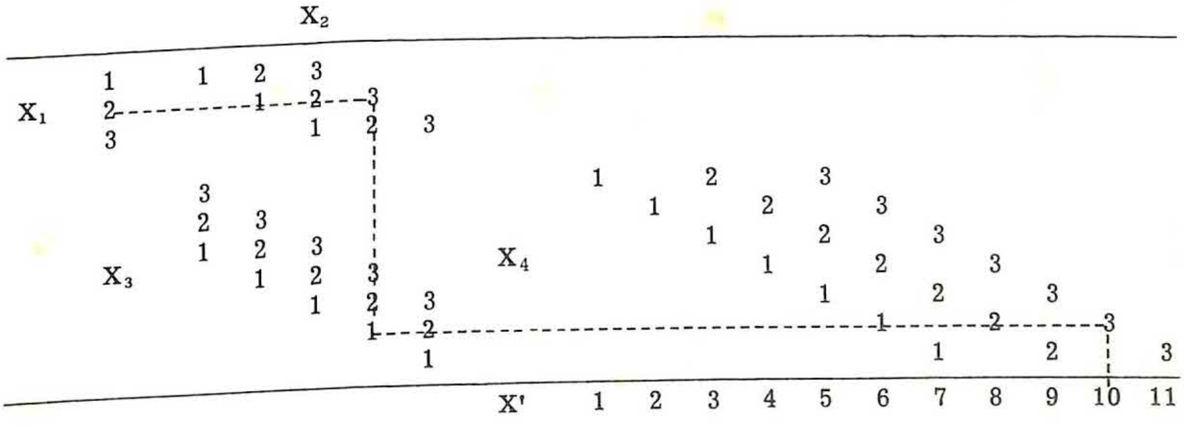
For predicting student grades, the following formula yielded a correlation of .69.

$$X' = .11X_1 + .09X_2 - .14X_3 + .26X_4 + 1.11$$

However, the weights are too complex to be used in a simple chart. Surprisingly, the following simplified form yielded a correlation of .68:

$$X' = X_1 + X_2 - X_3 + 2X_4$$

A chart was constructed where each variable had a range of ten. This proved quite convenient.



---

*Chicago Undergraduate Division.

The chart is presented here with a range of three, and with broken lines to indicate the path to the predicted score for a student with individual scores of 2, 3, 1, and 3.

It is preferrable, of course, to construct such a chart on paper with criss-cross lines so that the rows and columns may easily be followed.

For those interested in the mathematical theory of nomographic charts, two references are given.

REFERENCES

1. Douglas, R. D., and Adams, D. P. Elements of Nomography (New York: McGraw-Hill Book Co., 1947).
2. Mavis, F. T. The Construction of Nomographic Charts (Philadelphia: International Textbook Co., 1939).

Publishers: File two copies of this Form with your postmaster.

STATEMENT REQUIRED BY THE ACT OF AUGUST 24, 1912, AS AMENDED BY THE ACTS OF MARCH 3, 1933, JULY 2, 1946 AND JUNE 11, 1960 (74 STAT. 208) SHOWING THE OWNERSHIP, MANAGEMENT, AND CIRCULATION OF

Journal of Experimental Education (Insert exact title of publication) published Quarterly (State exact frequency of issue)

at Madison, Wisconsin (Name of post office and State where publication has second-class entry) for October 1, 1960

1. The names and addresses of the publisher, editor, managing editor, and business managers are:

	Name	*Address*
Publisher	Dembar Publications, Inc.	Madison, Wisconsin
Editor	A. S. Barr	Madison, Wisconsin
Managing editor	Jean W. Williams	Madison, Wisconsin
Business manager	Walter A. Frautschi	Madison, Wisconsin

2. The owner is: (If owned by a corporation, its name and address must be stated and also immediately thereunder the names and addresses of stockholders owning or holding 1 percent or more of total amount of stock. If not owned by a corporation, the names and addresses of the individual owners must be given. If owned by a partnership or other unincorporated firm, its name and address, as well as that of each individual member, must be given.)

Name	*Address*
Dembar Publications, Inc.	Madison, Wisconsin
A. S. Barr	Madison, Wisconsin
Walter A. Frautschi	Madison, Wisconsin
Dorothy J. Frautschi	Madison, Wisconsin

3. The known bondholders, mortgagees, and other security holders owning or holding 1 percent or more of total amount of bonds, mortgages, or other securities are: (If there are none, so state.)

Name	*Address*
None	

4. Paragraphs 2 and 3 include, in cases where the stockholder or security holder appears upon the books of the company as trustee or in any other fiduciary relation, the name of the person or corporation for whom such trustee is acting; also the statements in the two paragraphs show the affiant's full knowledge and belief as to the circumstances and conditions under which stockholders and security holders who do not appear upon the books of the company as trustees, hold stock and securities in a capacity other than that of a bona fide owner.

5. The average number of copies of each issue of this publication sold or distributed, through the mails or otherwise, to paid subscribers during the 12 months preceding the date shown above was: (This information is required by the act of June 11, 1960 to be included in all statements regardless of frequency of issue.) 927

Walter A. Frautschi Secy-Treas
(Signature of editor, publisher, business manager, or owner)

Sworn to and subscribed before me this 4th day of October, 1960

Lorna Canfield

[SEAL]

(My commission expires July 26, 1964)

POD Form 3526
(July 1960)

The Journal of Experimental Education

PUBLISHED QUARTERLY

Copyright 1961 by
DEMBAR PUBLICATIONS, INC.
Madison 3, Wisconsin

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 3, March 1961)

TEACHER VARIABLES OF WARMTH, DEMAND AND UTILIZATION OF INTRINSIC MOTIVATION RELATED TO PUPILS' SCIENCE INTERESTS: A STUDY ILLUSTRATING SEVERAL POTENTIALS OF VARIANCE-COVARIANCE*

HORACE B. REED, Jr.
Skidmore College
Saratoga Springs, New York

Purpose

THE PURPOSE of this article is to illustrate the variety of meaningful questions that can be asked concerning the data of a single research, and that can be dealt with largely through the tools of analysis of variance and covariance.

The research falls within the category of teacher competence studies, and is designed with the purpose of identifying selected teacher behaviors that relate to desirable pupil learning.

The Variables and Rationale

What is Competence?—The phrase "teacher competence" suggests the question, "Competent to do what?" The teacher's role as mediator of learning has been selected here to identify competence. This role has historical precedence, and most educational researchers now agree that pupil gains in learning are the most valid criteria of a teacher's competence (4, 5, 9, 28).

The Criterion Variable.—Selection of desirable pupil changes as the criteria in teacher competence research poses many difficulties. Present school objectives involve changes in pupils such as increased emotional maturity, ability to solve complex and realistic life problems, increased sensitivity to the welfare of others, enjoyment of intellectual and artistic pursuits, etc. Educators frankly have no adequate measures for such objectives. More limited objectives which may or may not lead to the "ultimate" school purposes are being measured, but not with universal confidence. At this level of abstraction would be found school goals such as pupils' accomplishment on subject matter tests, attitude tests, interest tests, art appreciation tests, comprehension tests, etc. A compromise with the direct measurement of learning has been suggested by Cogan (10). He has used pupils' performance of required school work and of self-initiated work as criteria, describing these as being proximate to what is ordinarily referred to as pupil change. Cogan's intervening criteria are an ingenious method of avoiding the dilemma of over-particularized measures (such as rate of eyelid conditioning) and over-generalized measures (such as asking pupils "From which teacher did you gain the most?").

In the present research, the criterion variable is interest in science, as measured by the author's Science Interest Inventory, a 70-item scale of the pupil's voluntary activities in science during the current school year. A number of factors have contributed to the selection of interest in science as the criterion: unlike many other pupil change criteria, interest is relatively independent of ability differences, especially up to mid-adolescence (13, 33); an analysis of some of the origins of interest suggests it is amenable to influences such as the teacher's classroom behaviors; it is sufficiently permanent to be an important predictor of many future activities of pupils; and interest is a worthy school objective in its own right.

Taking all these factors into account, the criterion of interest in the subject seems to meet a number of the common criticisms concerning the use of pupil change as a criterion in teacher competence studies.

Teacher Behaviors as Antecedents.—That the behavior of teachers produces a difference in the atmosphere of the classroom has been a truism for hundreds of years. Clear quantitative evidence for this is of relatively recent origin, perhaps not more than thirty years. The research on teachers' classroom personalities by Anderson and Brewer

---

*All footnotes will be found at end of article.

(3), and the research on leadership and group life headed by Kurt Lewin (24) demonstrated clearly that children behave differently while under the training of different "types" of adults. Withall (37) also demonstrated the existence of different psychological climates for different teachers with the same group of children. Thelen (34), Callis (8), Reed (31), Amatora (2), Bush (7), and Cogan (10), each using different measures, different samples, and different variables but all with careful designs, found unequivocal evidence that the differences among teachers can be quantitatively studied, that classroom atmosphere is partly a function of the teacher's behaviors, and that the teacher's behaviors measurably affect the pupils.

The teacher variables that may affect pupil learning can be categorized as either environmental or psychological. The first category would involve measures of the teacher's home background, number of years of schooling, the parents' religious preference, etc. It is reasonable to assume that such environmental factors (past and present) have an effect upon teacher behaviors. The second category would be concerned with the physical, emotional, and cognitive aspects of the teacher. The antecedent variables of this research fall within this latter category.

The Teacher Behavior Variable of Warmth—Warmth refers to pupils' perceptions[1] of teacher behaviors which relax interpersonal tension between teacher and pupil. There is a consequent reduction in the frequency of pupils' problems concerning "getting along with the teacher," but not necessarily in the frequency of problems concerning subject-matter tasks. Terms frequently used as synonyms of warmth are affection, affiliation, consideration, kindness, friendliness, sympathy, responsiveness, and geniality. Teacher behaviors expressive of the pupils' perceptions of warmth are illustrated by this item from the antecedent scales:

47. This teacher is careful about the feelings of pupils in this class
1-Not at all. 2-A little.
3-Somewhat. 4-Much.
5-Very much.

An hypothesis predicting a positive relationship between warmth and pupil interest in science is deduced from the following rationale.[2] The teacher's warm relationship is a rewarding experience for the pupils, classroom learning activities become rewarding as a function of the teacher's warmth behaviors. Pupils' positive feelings toward the learning activities lead to participation, which in turn, frequently leads to satisfaction of other needs such as the cognitive, as the learning activities become inherently attractive, interest is learned.

The Teacher Behavior Variable of Demand—Demand refers to the standards the teacher sets for each pupil's performance on school tasks. Demand is measured by items which refer to the pupil's perceptions of the teacher's level of demand as regards quantity, promptness, correctness, neatness, depth, thoroughness, honesty, attention, and orderliness. A sample item expressive of demand is:

15. This teacher demands that we do our science reading carefully
1-Not nearly strong enough.
2-Almost strongly enough.
3-Just strongly enough.
4-Somewhat too strongly.
5-Much too strongly.

An hypothesis that demand is non-linearly related to pupils' interest in science, with the relationship expressed as an inverted U-shape (moderate demand being the most effective), is based on the following ideas. Varying levels of demand generate within each pupil concomitant variations in tensions. This tension over school work may often function as an extrinsic motivation. The tensions can become so high that the pupil is unable to concentrate on the task, his energies are drained in many different directions in an effort to reduce the tensions to a bearable level. When tensions are very low, the pupil is not motivated by the challenge of utilizing his full abilities. When a moderate level of tension is established by the teacher's demands, the pupil is motivated to perform near the limits of his ability and/or willingness. In this situation, frequent success results in the schoolwork becoming enjoyable for its own sake.

The Teacher Behavior Variable of Utilization of Intrinsic Motivation—Intrinsic motivation refers to those motivations which the pupils have internalized; the learning activities have become meaningful to the pupils, either as means or as ends. The underlying task of the teacher is to present the curriculum in such a way that the pupil appreciates the relationships between it and his needs, interests, attitudes, and purposes. The following item is expressive of the teacher's utilization of intrinsic motivation:

26. When we start new work, this teacher helps us see why it is important to us
1-Almost never. 2-Few times
3-Sometimes. 4-Often.
5-Very often.

An hypothesis that intrinsic motivation is positively related to pupils' interest in science is based on the principle that if a teacher is successful in relating the curriculum to the pupils' needs, participation in the learning activities is inherently rewarding and a continuous feedback of energy is directed toward further voluntary involvement.

Design of the Research

The Control Variables—The science teacher's behaviors are not, of course, the only determinant of the pupils' science interest. The present research design employs four control variables in an attempt to reduce unaccounted-for variance in the criterion scores: sex of pupil, father's interest in science, school subject, and grade level.

Consistently, researches on science interests have demonstrated that boys' interest in science is stronger than girls'. The present analysis includes a study of classes (teacher-groups), and since the boy-girl ratio varies from class to class, mean class scores will reflect this variation. The effect of the sex variable will be tested and, if found significant, will be controlled through statistical procedures.

It is generally assumed that the interests of the father will have an influence on the interests of the children. As a simple index of the pupil's perceptions of his (her) father's interest in science, this item was included in the study:

> 50. How much interest does your father have in science?
> 1-Almost none. 2-A little.
> 3-Some. 4-Much. 5-Very much.

If a fairly strong relationship between father's science interest and pupil's science interest is confirmed by statistical analysis, and if classes differ significantly on mean father's science interest scores, differences among classes will be controlled through appropriate statistical procedures.

It is conjectured that the effects of teacher behaviors on pupil interest may vary as a function of the nature and objectives of the curriculum. Therefore, teachers and pupils of only one subject (science[3]) are used in the sample.

It is also conjectured that the effects of teacher behaviors may vary as a function of the developmental stage of the pupils. Therefore, pupils of only one grade (ninth) are used in the sample.

The Pupil Inventory—The Pupil Inventory is composed of two parts. Part I is the Science Interest Inventory, containing 70 science activity items. Each item is scored on a frequency scale from zero to five. Pupils were carefully instructed that only voluntary science activities performed during the present school year were to be reported. Voluntary science participation would appear to be a more valid measure of the pupils' science interest than would statements of likes and dislikes. Since only activities of the present school year were noted, the scores are a measure of performance during the same months that the pupils were with their present science teacher. This temporal concomitance of measured science interests of pupils and measured behaviors of science teachers increases the tenability of reasoning that a significant statistical relationship between these two variables also implies a degree of causal relationship.

Part II of the Pupil Inventory is composed of a random ordering of teacher behavior items; these provide scores for the three antecedent variables of warmth (13 items), demand (14 items), and utilization of intrinsic motivation (15 items). Each item is scored on an intensity or frequency scale from 1 to 5. The ascription of each item to a particular variable had been agreed upon by at least 90% of 30 scholars in the fields of education and psychology.

Administration of the Pupil Inventory requires less than 40 minutes.

The Sample—The sample used in the analysis included 1045 ninth-grade boys and girls and their 38 general science teachers, from 19 public schools representing 12 school systems in eastern Massachusetts. One class for each teacher was used, except that in three instances, two classes per teacher were inventoried and combined, because of the small size of the classes.

Administering the Pupil Inventory—The Pupil Inventory was administered by three trained assistants and the author. All data was collected in late May and early June of 1958, to insure that pupils were familiar with their science teacher's behaviors, and to maximize the teacher's influence on the science interests of the pupils. All teachers left the room during the administering of the Pupil Inventory, with the exception of two who worked at their desks because the principal requested they remain, on the basis of legal requirements.

Each administrator read printed instructions aloud to the pupils, emphasizing the privacy of the pupils' answers, the importance of honest and careful responses, and the serious nature of the research. A high level of pupils' concentration was observed.

Scoring the Pupil Inventories—Scoring procedures revealed six cases of pupils' apparent carelessness or failure to complete the Inventory adequately. Eleven other responses were not tabulated because the pupils had been in the science class for less than four months.

The 120 item responses on each of the remaining 1045 Pupil Inventories were then punched onto IBM cards. To verify the copying operation involved in card-punching, a print-out from the cards was proofread against the item responses in the original Inventories. New cards were punched for any in which errors were found.

Symbols Frequently Used in the Analysis

W The independent variable of the pupils' perceptions of the teacher's production of inter-

	personal warmth between teacher and pupil.
M	The independent variable of the pupils' perceptions of the extent of the teacher's utilization of intrinsic motivation.
D	The independent variable of the pupils' perceptions of the teacher's level of demand for pupils' performance on school tasks.
C	The independent variable for a compound measure of aspects of W and D.[4]
F	The independent variable of the pupil's perceptions of his (her) father's interest in science.
I	The dependent variable of the pupils' interest in science.
r	Subscript referring to a class (teacher-group).
i	Subscript referring to an individual score.
t, w, b	Subscripts referring to total scores, to within-classes scores, and to between-classes scores.
A	Sum of products of a dependent and an independent variable, with appropriate subscripts.
B	Sum of squares of a dependent variable, with appropriate subscripts.
C	Sum of squares of an independent variable, with appropriate subscripts.
N	Number of pupils in the sample.
k	Number of classes in the sample.

Questions, Statistical Resolutions, and Interpretations

What is the Reliability of the Scales for the Teacher Variables of W, M, and D?—For the scales measuring the teacher behavior variables, a reliability based on the stability of the within-classes responses of pupils is meaningful. Such a measure can be derived from the formula

$$r_{xx} = 1 - \frac{\sigma_w^2}{\sigma_b^2}$$

The values in Table I are taken from the analysis of variance-covariance reported later in this article. The reliabilities are between .88 and .93 for variables W and M, indicating that pupils of a given class agree fairly closely in their perceptions of these teacher behaviors. The reliabilities for D are somewhat lower (.78 to .80). A reasonable explanation for the lower reliability of the demand scale is that, as some teachers fail to vary their demands to meet pupil differences, variations in pupil abilities affect the pupils' perceptions of the level of a teacher's demands.

Can Single-Pupil Perceptions be Considered a Reliable Source Concerning Teacher Behaviors?—The standard error of measurement of the scales can be determined by calculating the square root of the within-classes variance estimate. With all three scales, the high standard errors of measurement listed in Table I signify that no single pupil's rating of his teacher is a reliable estimate of the score that the class will give that teacher. This does not mean that the perception of any given pupil is unreliable for that pupil. A pupil who perceives his science teacher as highly motivating, while most of his classmates rate the teacher low, will be affected by the teacher as he sees the teacher. The classroom science teacher can expect pupils to agree in general as to how they see him; he can also expect some divergence of individual pupil impressions. The concept of "a class" can lead to erroneous conclusions if the teacher supposes that his behaviors are uniformly perceived.

Are There Significant Differences Among the Teachers for the Antecedent Variables?—Table I shows the variance estimates for testing the statistical hypothesis of no difference among class means for each of the major variables of W, M, and D (data is also included for the criterion variable I). Acceptance of the hypothesis of no difference for any variable would render meaningless a between-classes analysis involving the variable. Since all F's are significant at the .001 level, this hypothesis is rejected. Each of the scales, then, did measure differences among teachers.

Do Boys and Girls Perceive Teachers Similarly as Regards the Teacher Variables of W, M, and D?—The fairly close agreement between the boy and girl means for W, M, and D shown in Table II suggests that the two sexes tend to rate their teachers in a similar fashion. Tests of the difference between boys' and girls' means for each of the three variables provided no significant CR's at the .01 level.

What is the Reliability of the Science Interest Inventory?—A reliability based on the stability of the within-classes responses of pupils is not applicable to the criterion scale measuring the individual pupil's interest in science; instead, an internal reliability concerned with the consistency of the items in the science activity scale is needed. Using a split-half reliability technique, the 70 items of the Science Interest Inventory were logically divided into two equivalent half-tests. The responses of the 1045 pupils were used for

TABLE I

VARIANCE ESTIMATES FOR I, W, M, AND D SCORES—BOYS (N = 584) AND GIRLS (N = 452)*

Source			Sum of Squares	df	Variance Estimate	r_{xx}	Standard Error	Variance Ratio**
I	Boys	Between	269,876	37	7293.94			2.18
		Within	1,825,848	546	3344.04			
	Girls	Between	196,603	33	5957.67			2.54
		Within	979,399	418	2343.06			
W	Boys	Between	27,863	37	753.05	.88	9.5	8.40
		Within	48,959	546	89.67			
	Girls	Between	41,408	33	1254.79	.93	9.4	14.16
		Within	37,034	418	88.60			
M	Boys	Between	24,703	37	667.65	.88	8.9	8.51
		Within	43,853	546	78.49			
	Girls	Between	25,188	33	763.27	.89	9.1	9.15
		Within	34,865	418	83.41			
D	Boys	Between	13,806	37	373.14	.78	9.1	4.53
		Within	45,005	546	82.43			
	Girls	Between	14,958	33	453.27	.80	9.6	4.90
		Within	38,651	418	92.47			

* This N for the girl sample is 9 less than the original N, as a result of the elimination of 4 classes with fewer than 5 girls per class. The total number of girl classes is 34, while the boy classes remain at 38.

**All F's are significant at .001 level.

TABLE II

COMPARISON OF THE TOTAL BOY AND THE TOTAL GIRL MEANS FOR W, M, AND D

Variable		Mean	N	S.D.	CR*
W	Boys	42.4	584	11.5	.77
	Girls	43.0	452	13.2	
M	Boys	42.9	584	10.8	.86
	Girls	42.3	452	11.5	
D	Boys	37.1	584	10.0	2.12
	Girls	35.7	452	10.9	

*No CR's are significant at the .01 level.

correlating the split-halves scores, and when the correction is made with the Spearman-Brown formula, a reliability coefficient of .97 is obtained for the whole criterion scale. This high value indicates internal consistency of the science interest items, and consistency of each pupil's responses.

Do Boys and Girls Have Equally Strong Science Interests?—In the original treatment of the data, this test was the first order of business, since a significant difference in boy and girl $\overline{I}$ scores would necessitate control of the sex variable. The $\overline{I}$ score of the 584 boys is 130.3; the $\overline{I}$ score for the 461 girls is 103.8. From Table III a variance ratio of 57.0 is obtained, and the statistical hypothesis of no difference is rejected at the .001 level.

TABLE III

VARIANCE ESTIMATES FOR TESTING SEX DIFFERENCE IN MEAN I SCORES

Source	Sum of Squares	df	Variance Estimate	Var. Ratio*
Between	181,637	1	181,637	57.0
Within	3,322,353	1043	3,185	

*Significant at the .001 level.

The difference between the two mean scores is so large that pooling of girl and boy scores would lead to major distortion in the analysis involving classes. Therefore, the total sample is separated into two sub-samples, one for the boys and one for the girls. At this point the total number of girls is reduced from 461 to 452, after elimination of four classes with fewer than five girls per class. The number of boy classes is 38; the number of girl classes is 34.

Is There a Significant Relationship Between Father's Interest in Science and Pupil's Interest in Science?—Table IV contains data on the product-moment correlation coefficient for F and I. The obtained r_{FI} for boys (.40) and r_{FI} for girls (.27) are based on total samples. Application of the significance test indicates both r's to be significant at the .001 level.

TABLE IV

RELATIONSHIP OF F AND I

Sample	r_{FI}	Variance Ratio*	df
Total, Boys (N=584)	.40	108.4	$n_1 = 1$ $n_2 = 582$
Total, Girls (N=452)	.27	34.4	$n_1 = 1$ $n_2 = 450$

*Both ratios are significant at .001 level.

In Studying Teacher Effects, Does the Influence of F Upon I Need to be Controlled?—The obtained significant relationships of r_{FI} pose the problem of whether there is sufficient attenuation of $\overline{I}$ scores for classes, attributable to differences among $\overline{F}$ scores for classes, to weigh in the choice of methods of analysis. Table V presents the requisite variance estimates for the test of significant differences among the boys' $\overline{F}$ scores for classes, and also among the girls' $\overline{F}$ scores for classes. The obtained variance ratios for boys of 1.56, and for girls of 1.62, are not significant at the .01 level, but are at the .05. The statistical hypothesis of no differences among $\overline{F}$ scores for classes is accepted, with some reservations.

TABLE V

VARIANCE ESTIMATES FOR TESTING DIFFERENCES AMONG $\overline{F}$ SCORES FOR CLASSES

Source	Sum of Squares	df	Variance Estimate	Var. Ratio*
Between, Boys	78.0	37	2.10	1.56
Within, Boys	736.9	546	1.35	
Between, Girls	64.7	33	1.96	1.62
Within, Girls	506.5	418	1.21	

*Neither ratio is significant at the .01 level, but both are significant at .05.

McNemar (27:354), in discussing procedures for adjusting scores by the covariance method, writes:

> . . . . But, if the within-group correlation is low and/or there is only a small chance difference between the groups on the uncontrolled variable, the use of the covariance adjustment may not be worth the effort. . . .

The moderate strength of the relationship of F and I, and the lack of significant differences at the .01 level among $\overline{F}$ scores for classes, demonstrates that a very limited amount of the variability of $\overline{I}$ scores for classes can be reduced by correcting for differences due to F. Therefore, the attenuation of $\overline{I}$ scores for classes, due to the influence of F, was considered insufficient to dictate subsequent methods of analysis.

Does the Father's Science Interest Have a Greater Influence on Son's Interest Than on Daughter's?—A test of any significant difference in the r_{FI} for the boy sample (.40) and the girl sample (.27) can be made using Fisher's r to z transformation. The resulting critical ratio of 2.33 is significant at the .02 level. The statistical hypothesis of no significant difference is rejected at the .02 level. The demonstration of this difference supports the reasoning that, at the ninth-grade level, sons tend to imitate their fathers somewhat more closely than do daughters.

Is There a Significant Relationship Between D and I, and is it Non-Linear?—Values for the linear correlation and for the correlation ratio are in Table VI. The relevant variance estimates for determining eta, and the tests for significance of eta and of linearity of regression, are shown in Table VII. The ratio of variance estimates for between-array means and for within arrays provides a test for the significance of eta. The obtained F's of .81 for the boy sample and 1.43 for the girl sample indicate large within-arrays variance. For both the boy and the girl samples, the linear relationship and the correlation ratio are not significant at the .01 level. A further test for significance of curvilinearity would be meaningless.

TABLE VI

RELATIONSHIP OF D AND I

Type of Relationship	Strength of Relationship*	
	Boys (N=584)	Girls (N=461)**
Linear (r)	.04	.04
Correlation Ratio (eta)	.13	.19

* None of the relationships is significant at the .01 level.

**This analysis, based on the coded scores of the total samples, was made prior to the major analysis involving raw scores and computer techniques. At this stage, the reduction of the girls' N to 452 had not yet been made.

While the predicted positive effect of mild tension was not demonstrated, it is important to emphasize that this research did not demonstrate the negative relationship between tension and pupil gain traditionally hypothesized. This consistent lack of a relationship between teacher demand and the broad school objective of pupils' interest in science may encourage other researchers to investigate the relationship of teacher demand to more specific criteria, such as subject matter achievement. There is the evidence from this research that the more general goal of interest will not necessarily be adversely affected as a concomitant of the teacher's use of demand.

What is the Relationship Between M and I?—In this study, the strongest relationship between any teacher behavior variable and the criterion is between M and I. The most accurate measure of the relationship based on individual pupil scores is provided by the within-classes r_{MI}, which corrects for possible spuriousness evidenced by between-classes correlation. Tables VIII and IX show the within-classes r_{MI} values: .32 for boys and .42 for girls, both significant at the .001 level. Also from Tables VIII and IX, the between-classes r_{MI} for boys (.40) is significant at the .05 level; the between classes r_{MI} for girls (.53) is significant at the .01 level.

There is no significant difference at the .01 level between these correlations of M and I for the boy sample and the comparable correlations of M and I for the girl sample.

For the form of the relationship, a test of the hypothesis of linear regression for the total r_{MI} can be made through the formula

$$F = \frac{S_d^2}{S_w^2}$$

k = number of arrays
n_1 = k - 2
n_2 = N - k

where S_d^2 represents the variance estimate based on the deviation of array means from the line and S_w^2 represents the variance estimate based on the deviation of individual scores within arrays. Values for the F ratios of the two samples are in Table X. Since both F's are less than unity, the hypothesis of linear regression is accepted for each sample.

Two of the variance estimates of Tables XI and XII may be used to provide an F ratio to test for linearity of regression based on the means of M and I (between-classes r_{MI}):

TABLE VII

VARIANCE ESTIMATES FOR STUDYING THE FORM OF r_{DI}

Source	Boys (N = 584)			Girls (N = 461)		
	Sum of Squares	df	Variance Estimate	Sum of Squares	df	Variance Estimate
Linear Regression	7.85	1	7.85	5.30	1	5.30
Deviation of Means from Line	66.31	11	6.03	110.14	11	10.01
Between-Array Means	72.60	12	6.05	114.86	12	9.57
Within Arrays	4290.10	571	7.51	3005.34	448	6.71
Residual from Line	4354.85	582	7.48	3114.90	459	6.79

TABLE VIII

ANALYSIS OF VARIANCE:COVARIANCE, FOR FUNCTIONS OF M AND I—BOYS

	Total	Within	Between
Sum of Products M and I	$A_t = 123,118$	$A_w = 90,263$	$A_b = 32,855$
Sum of Squares, I	$B_t = 2,095,724$	$B_w = 1,825,848$	$B_b = 269,876$
Sum of Squares, M	$C_t = 67,556$	$C_w = 42,853$	$C_b = 24,703$
df for Variances	$N - 1 = 583$	$N - k = 546$	$k - 1 = 37$
Correlation Coefficient, r_{MI}	$\frac{A_t}{\sqrt{B_t}\sqrt{C_t}} = .33$	$\frac{A_w}{\sqrt{B_w}\sqrt{C_w}} = .32$	$\frac{A_b}{\sqrt{B_b}\sqrt{C_b}} = .40$
df for r_{MI}	$N - 2 = 582$	$N - k - 1 = 545$	$k - 2 = 36$
Significance Level for r_{MI}	Sig. at .001 level	Sig. at .001 level	Sig. at .05 level
Regression Coefficient, b_{IM}	$\frac{A_t}{C_t} = 1.82$	$\frac{A_w}{C_w} = 2.11$	$\frac{A_b}{C_b} = 1.33$

TABLE IX

ANALYSIS OF VARIANCE:COVARIANCE, FOR FUNCTIONS OF M AND I—GIRLS

	Total	Within	Between
Sum of Products M and I	$A_t = 114,954$	$A_w = 77,731$	$A_b = 37,223$
Sum of Squares, I	$B_t = 1,176,002$	$B_w = 979,399$	$B_b = 196,603$
Sum of Squares, M	$C_t = 60,053$	$C_w = 34,865$	$C_b = 25,188$
df for Variances	$N - 1 = 451$	$N - k = 418$	$k - 1 = 33$
Correlation Coefficient, r_{MI}	$\frac{A_t}{\sqrt{B_t}\sqrt{C_t}} = .43$	$\frac{A_w}{\sqrt{B_w}\sqrt{C_w}} = .42$	$\frac{A_b}{\sqrt{B_b}\sqrt{C_b}} = .53$
df for r_{MI}	$N - 2 = 450$	$N - k - 1 = 417$	$k - 2 = 32$
Significance Level for r_{MI}	Sig. at .001 level	Sig. at .001 level	Sig. at .01 level
Regression Coefficient, b_{IM}	$\frac{A_t}{C_t} = 1.91$	$\frac{A_w}{C_w} = 2.23$	$\frac{A_b}{C_b} = 1.48$

TABLE X

VARIANCE ESTIMATES FOR STUDYING THE FORM OF r_{MI}

Source	Boys (N = 584) Sum of Squares	Boys (N = 584) df	Boys (N = 584) Variance Estimates	Girls (N = 452) Sum of Squares	Girls (N = 452) df	Girls (N = 452) Variance Estimates
Linear Regression	419. 26	1	419. 26	433. 06	1	433. 06
Deviation of Means from Line	60. 64	12	5. 05	33. 39	12	2. 78
Between-Array Means	460. 45	13	35. 42	472. 30	13	36. 33
Within Arrays	3902. 25	570	6. 85	1982. 30	438	4. 53
Residual from Line	3943. 44	582	6. 78	2021. 94	450	4. 49

TABLE XI

VARIANCE ESTIMATES FOR TESTS OF SLOPES b_{IM}—BOYS

Source of Variance	Formula and Variance	df	Variance Estimate
Class r's About Common Coefficient	$\sum_{1}^{k} \frac{A_{ir}^2}{C_{ir}} - \frac{A_w^2}{C_w} =$ 61,881	k - 1 = 37	$S_1^2 = 1672.5$
Individual Scores About Regression Line of Each Class	$B_w - \sum_{1}^{k} \frac{A_{ir}^2}{C_{ir}}$ = 1,573,842	N - 2k = 508	$S_2^2 = 3098.1$
Group Means About Regression Line based on Means	$B_b - \frac{A_b^2}{C_b}$ = 226,179	k - 2 = 36	$S_3^2 = 6282.8$
Individual Scores About Regression Line of Common Within Slope	$B_w - \frac{A_w^2}{C_w}$ = 1,635,723	N - k - 1 = 545	$S_w^2 = 3001.3$
Group Means About Regression Line of Common Within Slope	$B_b + \frac{A_w^2}{C_w} - \frac{A_t^2}{C_t} =$ 235,624	k - 1 = 37	$S_b^2 = 6368.2$
Individual Scores About Regression Line for Total Group	$B_t - \frac{A_t^2}{C_t}$ = 1,871,347	N - 2 = 582	

TABLE XII

VARIANCE ESTIMATES FOR TESTS OF SLOPES b_{IM}—GIRLS

Source of Variance	Formula and Variance	df	Variance Estimate
Class r's About Common Coefficient	$\sum_{1}^{k} \frac{A_{ir}^2}{C_{ir}} - \frac{A_w^2}{C_w} = 39,801$	$k - 1 = 33$	$S_1^2 = 1206.1$
Individual Scores About Regression Line of Each Class	$B_w - \sum_{1}^{k} \frac{A_{ir}^2}{C_{ir}} = 766,298$	$N - 2k = 384$	$S_2^2 = 1995.6$
Group Means About Regression Line Based on Means	$B_b - \frac{A_b^2}{C_b} = 141,595$	$k - 2 = 32$	$S_3^2 = 4424.8$
Individual Scores About Regression Line of Common Within Slope	$B_w - \frac{A_w^2}{C_w} = 806,099$	$N - k - 1 = 417$	$S_w^2 = 1933.1$
Group Means About Regression Line of Common Within Slope	$B_b + \frac{A_w^2}{C_w} - \frac{A_t^2}{C_t} = 149,857$	$k - 1 = 33$	$S_b^2 = 4541.1$
Individual Scores About Regression Line for Total Group	$B_t - \frac{A_t^2}{C_t} = 955,956$	$N - 2 = 450$	

$$F = \frac{S_3^{\,2}}{S_W^{\,2}} \qquad n_1 = k - 2 \quad n_2 = N - k - 1$$

The obtained variance ratios of 2.09 for boys and 2.29 for girls are significant at the .001 level. This demonstrates that the estimate of the variance of class means about the regression line based on means (S_3^2) is significantly larger than the estimate of the variance of individual scores about the regression line of common within-classes slope (S_W^2). Therefore, the hypothesis of linearity of regression based on means is rejected for both the boy and the girl sample.

Are the 38 Separate Class Slopes b_{IM} for Boys Equivalent, and are the 34 Separate Class Slopes for Girls Equivalent?—Tables XI and XII provide the necessary variance estimates to test the hypothesis that slope b_{IM} of the within-classes regression line is equivalent for all classes of a given sex sample. The formula

$$F = \frac{S_1^{\,2}}{S_2^{\,2}} \qquad n_1 = k - 1 \quad n_2 = N - 2k$$

results in an F of .54, for the boy sample. A similar test for the girl sample results in an F that is also less than unity. The estimate of the variance of class r's about the common coefficient (S_1^2) is less than the estimate of the variance of the individual scores about the regression line of each class (S_2^2). The hypothesis of a common within-classes slope for all boy classes and a common within-classes slope for all girl classes, is accepted. For boys, this common within-classes b_{IM} is 2.11; for girls, the common within-classes b_{IM} is 2.23 (Tables VIII and IX).

The consistency of the relationship of the teacher's utilization of intrinsic motivation and pupils' interest in science can be appreciated from the above demonstration that in each sample (for boys and for girls) the highly significant within-classes slope is common for all classes. That is, any variations among the 38 separate class slopes b_{IM} for boys, and among the 34 separate class slopes b_{IM} for girls, can be attributed to sampling error.

Figure 1 illustrates for the girl sample the regressions for the total slope, for the common within-classes slope, and for the slope of the class means. The variations[5] of the 34 separate class slopes from the common within-classes slope are shown. Also indicated is the scatter of the class means about the regression line determined by those means. The girl sample is selected for illustration because it provides one more highly significant regression (for between-classes) than was demonstrated for the boy sample. For simplicity, the slope for the class means is shown here as a linear regression, although it has been established that the actual regression is curvilinear.

The three regression equations, and values for the appropriate slopes, are summarized in Table XIII. The equation for the 34 separate class slopes is given; the separate values are too detailed to list.

Are There Factors, in Addition to M and Other Controlled Variables, Which are Affecting the Criterion Scores?—To resolve this question it is necessary to test the hypothesis that a single regression line fits both total and between-classes populations. To test this hypothesis, the following variance ratio is appropriate:

$$F = \frac{S_b^{\,2}}{S_W^{\,2}} \qquad n_1 = k - 1 \quad n_2 = N - k - 1$$

The relevant variance estimates for the two samples are presented in Tables XI and XII. The F ratios for boys (2.12) and for girls (2.10) are both significant at the .001 level. Therefore, the estimate of the variance of group means about the regression line of the common within-classes slope (S_b^2) is significantly larger than the estimate of the variance of individual scores about the regression line of common within-classes slope (S_W^2). The hypothesis of a common regression line for all populations is rejected for both samples, since acceptance would require that S_b^2 and S_W^2 be similar.

The lack of a common slope means that with respect to classes there are factors other than teachers' utilization of intrinsic motivation which account for some of the variation in interest scores among the classes. Certain of these attenuating factors had been anticipated prior to the collection of the data, and various methods of control were included in the design to eliminate or reduce their possible effects (sex of pupil, father's science interest, subject matter, and grade level). Many other possible forces were beyond the scope of this study, such as additional characteristics of the teachers, school atmosphere, and previous experience of the pupils.

What is the Relationship Between W and I?[6]—The product moment correlation coefficients for W and I are calculated from the information in Tables XIV and XV. The within-classes r_{WI} values are .20 for boys, and .28 for girls, both significant at the .001 level.

The between-classes r_{WI} for boys of .23 is not significant at the .01 level. The higher between-classes r_{WI} for girls of .53 is significant at the .01 level.

A test of the difference between the boy and the girl within-classes r_{WI}, using Fisher's r to z transformation, results in a CR of 1.35. This ratio is not significant at the .01 level. The hypothesis that the two r's are not significantly different is accepted. A test of the difference between the boy and the girl between-classes r_{WI} results in a CR of 1.44, which is not significant

FIGURE 1

RELATIONSHIP BETWEEN M AND I (GIRLS)

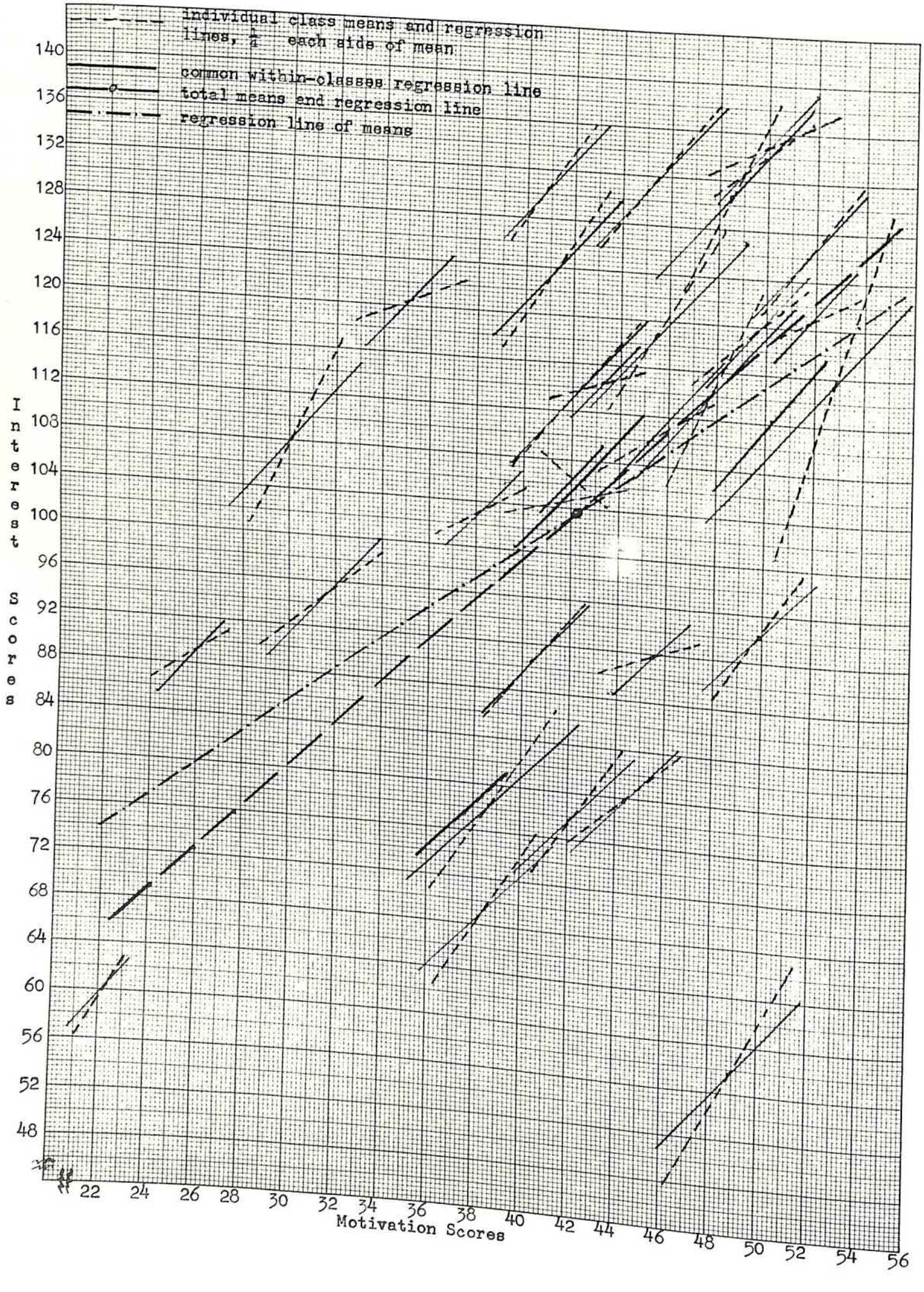


TABLE XIII

REGRESSION EQUATIONS FOR M AND I—GIRLS

Source of Regression	Equation	Slope
Total	$I_i' = \bar{I} + b_t(M_i - \bar{M})$	$b_t = 1.91$
Common Within-Classes	$I_{ir}' = \bar{I}_r + b_w(M_{ir} - \bar{M}_r)$	$b_w = 2.23$
Between Classes	$\bar{I}_r' = \bar{I} + b_b(\bar{M}_r - \bar{M})$	$b_b = 1.48$
Any Specific Class	$I_{ir}' = \bar{I}_r + b_r(M_{ir} - \bar{M}_r)$	b_r = (34 separate values)

TABLE XIV

ANALYSIS OF VARIANCE:COVARIANCE, FOR FUNCTIONS OF W AND I—BOYS

	Total	Within	Between
Sum of Products, W and I	$A_t = 80,727$	$A_w = 60,703$	$A_b = 20,024$
Sum of Squares, I	$B_t = 2,095,724$	$B_w = 1,825,848$	$B_b = 269,876$
Sum of Squares, W	$C_t = 76,822$	$C_w = 48,959$	$C_b = 27,863$
df for Variances	$N - 1 = 583$	$N - k = 546$	$k - 1 = 37$
Correlation Coefficient, r_{WI}	$\frac{A_t}{\sqrt{B_t}\sqrt{C_t}} = .20$	$\frac{A_w}{\sqrt{B_w}\sqrt{C_w}} = .20$	$\frac{A_b}{\sqrt{B_b}\sqrt{C_b}} = .23$
df for r_{WI}	$N - 2 = 582$	$N - k - 1 = 545$	$k - 2 = 36$
Significance Level for r_{WI}	Sig. at .001 level	Sig. at .001 level	Not sig. at .01 level

TABLE XV

ANALYSIS OF VARIANCE:COVARIANCE, FOR FUNCTIONS OF W AND I—GIRLS

	Total	Within	Between
Sum of Products, W and I	$A_t = 101,399$	$A_w = 53,622$	$A_b = 47,777$
Sum of Squares, I	$B_t = 1,176,002$	$B_w = 979,399$	$B_b = 196,603$
Sum of Squares, W	$C_t = 78,442$	$C_w = 37,034$	$C_b = 41,408$
df for Variances	$N - 1 = 451$	$N - k = 418$	$k - 1 = 33$
Correlation Coefficient, r_{WI}	$\frac{A_t}{\sqrt{B_t}\sqrt{C_t}} = .33$	$\frac{A_w}{\sqrt{B_w}\sqrt{C_w}} = .28$	$\frac{A_b}{\sqrt{B_b}\sqrt{C_b}} = .53$
df for r_{WI}	$N - 2 = 450$	$N - k - 1 = 417$	$k - 2 = 32$
Significance Level for r_{WI}	Sig. at .001 level	Sig. at .001 level	Sig. at .01 level

TABLE XVI

DATA FOR MULTIPLE R

Boys			Girls		
Betas	Zero Order r's	Product of B and r	Betas	Zero Order r's	Product of B and r
$W{:}B_2 = -.147$	$r_{WI} = .20$	-.029	$W{:}B_2 = -.024$	$r_{WI} = .33$	-.008
$M{:}B_3 = .335$	$r_{MI} = .33$	.111	$M{:}B_3 = .396$	$r_{MI} = .43$	.170
$D{:}B_4 = -.091$	$r_{DI} = .03$	-.003	$D{:}B_4 = -.096$	$r_{DI} = .03$	-.003
$C{:}B_5 = .107$	$r_{CI} = .25$	.027	$C{:}B_5 = .083$	$r_{CI} = .34$	.028
$F{:}B_6 = .362$	$r_{FI} = .40$	.145	$F{:}B_6 = .219$	$r_{FI} = .27$	.059

at the .01 level.

Can Prediction of I be Increased Through a Study of the Relative Contributions of W, M, D, and F?—The relative contribution of each of the antecedent variables to the prediction of the independent variable I can be observed in Table XVI (products of B's and r's). For both the boy and the girl samples, the two variables of M and F contribute the most weight, with F the more predictive for boys and M the more for girls.

The value of R (multiple correlation) is .50 for the boy and for the girl sample. The obtained F for boys is 37.9, and for girls is 29.2; both R's are significant at the .001 level.

While the many origins of pupils' interest in science do not constitute the central problem of this study, it is worth noting that 25% of the variance in pupils' science interest scores can be accounted for by a knowledge of the variation in scores for the teacher's utilization of intrinsic motivation and for the father's interest in science.

Does the Effect of One Teacher Variable Upon I Vary as a Function of Other Teacher Variables?—In Tables XVII and XVIII, four cells have been formed by a double classification based on class mean scores for high and low motivation, and for moderate and low demand (there were no high $\bar{D}$ scores for classes). The cell values are derived from the corresponding $\bar{I}_a$ scores for classes. The variations of these cell means suggests that the differences in $\bar{I}_a$ scores are due not only to the effects of high and low M, but also to some other influence. As D has no significant correlation with I, it is possible that some of the differences in the cell means are due to the interaction of M and D.

A test of the hypothesis of interaction of M and D is made through complex analysis of variance. Prior to this analysis, two steps were taken. First, to account for as much within-cell variability of $\bar{I}$ scores for classes as possible, each $\bar{I}$ score was adjusted for differences due to $\bar{F}$ scores for classes,[6] using the regression equation for adjusted scores:

$$\bar{I}_{ra} = \bar{I}_r - b_{IFw}(\bar{F}_r - \bar{F})$$

The reduction in variability can be partially gauged by comparing the unadjusted with the adjusted sum of squares for I, as summarized in Tables XIX and XX.

In the second prior step, correction for disproportionality of cell frequencies was made through procedures outlined by Wert, Neidt, and Ahmann (36). The adjustment term is added to the sum of squares for M and D, and is subtracted from the sum of squares for interaction.

Table XXI summarizes the interaction data. The variance ratios to test the main effects and to test interaction are formed by using the within variance estimate as the denominator term. As indicated by previous correlation analysis of $\bar{M}$ and $\bar{I}$ scores for classes, the main effect of motivation is significant for girls at the .01 level, and for boys at the .05 level. In both samples, the non-significant variance ratio for the main effect of demand is also to be expected from previous analysis. Since neither variance ratio for interaction is significant at the .01 level, the hypothesis of no interaction of $\bar{M}$ scores for classes with $\bar{D}$ scores for classes is accepted for both samples. If the interaction effect does exist, it is not sufficiently strong to be demonstrated with the number of classes available for study.

Is There an Interaction Effect of M and D, Using Pupil I Scores as Criterion?—Increased cell frequencies may be obtained by using pupil scores as a basis for further study of interaction of M with D, using cells formed essentially as in Tables XVII and XVIII. Individual I scores are not adjusted for differences due to the influence of F, since control by covariance adjustment is appropriate only with groups. Adjustment is made for disproportionality of cell frequencies. The variance estimates for this test of interaction are presented in Table XXII. For both samples, the variance ratio is significant at the .001 level for the main effect of M. The main effect of D is not significant at the .01 level for either sample. The variance ratio for interaction is less than unity in both samples, demonstrating large within-cell variance. The hypothesis of no interaction is accepted.

A similar series of tests for interaction of W and D, using pupil I scores as the criterion, resulted in acceptance of the hypothesis of no interaction for this combination of variables.

Does Sex of Teacher, Interacting with Sex of Pupil, have an Effect on Pupils' I Scores?—As the sample includes both women teachers (N = 10) and men teachers (N = 28), it is possible to test for differences in pupil I scores attributable to teacher sex. (The test of significance for the main effect of pupil sex has been discussed previously.) Table XXIII shows the variance estimates for testing interaction and the main effect of teacher sex. The hypothesis of no difference due to teacher sex is accepted, since the obtained variance ratio is less than unity. The hypothesis of no difference due to interaction is accepted, since the obtained F of 2.07 is not significant at the .01 level.

Does Sex of Teacher, Interacting with Sex of Pupil, have an Effect on Pupil Perceptions of Teacher Behaviors?—Tests for any differences in pupil perceptions of men and women teachers for D, M, and W can be made using the values in Table XXIV. (The lack of significant differences in perceptions of D, M, and W due to pupil sex has been previously discussed.) The obtained variance ratios for the main effects of teacher sex and for interaction, are not significant at the .01

TABLE XVII

CLASSIFICATION OF $\overline{I}_a$ SCORES FOR CLASSES,* BASED ON TWO LEVELS OF $\overline{M}$ SCORES FOR CLASSES AND TWO LEVELS OF $\overline{D}$ SCORES FOR CLASSES—BOYS (N = 38)

		Motivation	
		High	Low
Demand	Moderate	1 $\overline{I}_1$ = 145.2 11 cases	2 $\overline{I}_2$ = 122.3 8 cases
	Low	3 $\overline{I}_3$ = 130.7 8 cases	4 $\overline{I}_4$ = 127.4 11 cases

*The $\overline{I}$ scores for classes have been adjusted for class differences due to F.

TABLE XVIII

CLASSIFICATION OF $\overline{I}_a$ SCORES FOR CLASSES,* BASED ON TWO LEVELS OF $\overline{M}$ SCORES FOR CLASSES AND TWO LEVELS OF $\overline{D}$ SCORES FOR CLASSES—GIRLS (N = 34)

		Motivation	
		High	Low
Demand	Moderate	1 $\overline{I}_1$ = 111.9 10 cases	2 $\overline{I}_2$ = 83.7 7 cases
	Low	3 $\overline{I}_3$ = 114.4 7 cases	4 $\overline{I}_4$ = 104.8 10 cases

*The $\overline{I}$ scores for classes have been adjusted for class differences due to F.

TABLE XIX

ANALYSIS OF VARIANCE:COVARIANCE, FOR FUNCTIONS OF F AND I—BOYS

	Total	Within	Between
Sum of Products, F and I	A_t = 16,357	A_w = 13,842	A_b = 2,515
Sum of Squares, I	B_t = 2,095,724	B_w = 1,825,848	B_b = 269,876
Sum of Squares, F	C_t = 815	C_w = 737	C_b = 78
df for Variances	N - 1 = 583	N - k = 546	k - 1 = 37
Correlation Coefficient, r_{FI}	$\frac{A_t}{\sqrt{B_t}\sqrt{C_t}} = .40$	$\frac{A_w}{\sqrt{B_w}\sqrt{C_w}} = .38$	$\frac{A_b}{\sqrt{B_b}\sqrt{C_b}} = .55$
Regression Coefficient, b_{IF}	$\frac{A_t}{C_t} = 20.1$	$\frac{A_w}{C_w} = 18.8$	$\frac{A_b}{C_b} = 32.2$
Adjusted Sum of Squares, I	$B_t - \frac{A_t^2}{C_t}$ minus 1,767,360	$B_w - \frac{A_w^2}{C_w}$ equals 1,565,839	Adjusted B_b 201,521
df for Adjusted Sum of Squares	N - 2 = 582	N - k - 1 = 545	k - 1 = 37

TABLE XX

ANALYSIS OF VARIANCE:COVARIANCE, FOR FUNCTIONS OF F AND I—GIRLS

	Total	Within	Between
Sum of Products, F and I	A_t = 6,904	A_w = 5,712	A_b = 1,192
Sum of Squares, I	B_t = 1,176,002	B_w = 979,399	B_b = 196,603
Sum of Squares, F	C_t = 571	C_w = 506	C_b = 65
df for Variances	N - 1 = 451	N - k = 418	k - 1 = 33
Correlation Coefficient, r_{FI}	$\frac{A_t}{\sqrt{B_t}\sqrt{C_t}} = .27$	$\frac{A_w}{\sqrt{B_w}\sqrt{C_w}} = .26$	$\frac{A_b}{\sqrt{B_b}\sqrt{C_b}} = .70$
Regression Coefficient, b_{IF}	$\frac{A_t}{C_t} = 12.1$	$\frac{A_w}{C_w} = 11.3$	$\frac{A_b}{C_b} = 18.4$
Adjusted Sum of Squares, I	$B_t - \frac{A_t^2}{C_t}$ minus 1,092,554	$B_w - \frac{A_w^2}{C_w}$ equals 914,983	Adjusted B_b 177,571
df for Adjusted Sum of Squares	N - 2 = 450	N - k - 1 = 417	k - 1 = 33

TABLE XXI

VARIANCE ESTIMATES FOR TEST OF INTERACTION OF M WITH D SCORES FOR CLASSES—BOYS (N = 38) AND GIRLS (N = 34)

	Source	Adjusted Sum of Squares*	df	Variance Estimate	Variance Ratio	Level of Significance
Boys	Motivation	1882. 0	1	1882. 0	6. 15	Sig. at . 05 level
	Demand	492. 1	1	492. 1	1. 61	Not sig. at . 01 level
	Interaction	604. 4	1	604. 4	1. 98	Not sig. at . 01 level
	Within	10401. 0	34	305. 9		
Girls	Motivation	2950. 1	1	2950. 1	9. 50	Sig. at . 01 level
	Demand	1163. 6	1	1163. 6	3. 75	Not sig. at . 01 level
	Interaction	704. 5	1	704. 5	2. 27	Not sig. at . 01 level
	Within	9315. 5	30	310. 5		

*Sum of squares adjusted for disproportionality of cells.

TABLE XXII

VARIANCE ESTIMATES FOR TEST OF INTERACTION OF M WITH D SCORES—BOYS (N = 584) AND GIRLS (N = 452)

	Source	Adjusted Sum of Squares*	df	Variance Estimate	Variance Ratio	Level of Significance
Boys	Motivation	148, 054	1	148, 054	44. 14	Sig. at . 001 level
	Demand	11, 165	1	11, 165	3. 33	Not sig. at . 01 level
	Interaction	2, 110	1	2, 110	. 63	Not sig. at . 01 level
	Within	1, 945, 123	580	3, 354		
Girls	Motivation	123, 814	1	123, 814	53. 18	Sig. at . 001 level
	Demand	222	1	222	. 10	Not sig. at . 01 level
	Interaction	868	1	868	. 37	Not sig. at . 01 level
	Within	1, 043, 112	448	2, 328		

*Sum of squares adjusted for disproportionality of cells.

TABLE XXIII

VARIANCE ESTIMATES FOR TESTING EFFECTS OF TEACHER SEX ON BOY AND GIRL I SCORES (N = 1036)

Source	Adjusted Sum of Squares*	df	Variance Estimate	Variance Ratio	Level of Significance
Teacher Sex	1,303	1	1,303	.41	Not sig. at .01 level
Pupil Sex	182,902	1	182,902	58.83	Sig. at .001 level
Interaction	6,546	1	6,546	2.07	Not sig. at .01 level
Within	3,263,875	1032	3,163		

*Sum of squares adjusted for disproportionality of cells.

TABLE XXIV

VARIANCE ESTIMATES FOR TESTING EFFECTS OF TEACHER SEX ON BOY AND GIRL PERCEPTIONS OF D, M, AND W (N = 1036)

	Source	Adjusted Sum of Squares*	df	Variance Estimate	Variance Ratio	Level of Significance
D	Teacher Sex	314	1	314	2. 89	Not sig. at . 01 level
	Pupil Sex	488	1	488	4. 49	Not sig. at . 01 level
	Interaction	3	1	3	. 03	Not sig. at . 01 level
	Within	112, 103	1032	109		
M	Teacher Sex	297	1	297	2. 41	Not sig. at . 01 level
	Pupil Sex	112	1	112	. 91	Not sig. at . 01 level
	Interaction	204	1	204	1. 66	Not sig. at . 01 level
	Within	127, 108	1032	123		
W	Teacher Sex	5, 959	1	5959	41. 30	Sig. at . 001 level
	Pupil Sex	41	1	41	. 28	Not sig. at . 01 level
	Interaction	409	1	409	2. 84	Not sig. at . 01 level
	Within	148, 896	1032	144		

*Sum of squares adjusted for disproportionality of cells.

TABLE XXV

CLASSIFICATION OF W SCORES, BASED ON TEACHER SEX AND PUPIL SEX (N = 1036)

		Teacher Sex	
		Male	Female
Pupil Sex	Boy	1 $\overline{W}_1 = 43.50$ 439 cases	2 $\overline{W}_2 = 39.12$ 145 cases
	Girl	3 $\overline{W}_3 = 44.61$ 351 cases	4 $\overline{W} = 37.23$ 101 cases

level with one important exception: the main effect of teacher sex with W as the criterion. The obtained variance ratio of 41.30 for differences in pupils' perceptions of W attributable to teacher sex is significant at the .001 level. Table XXV demonstrates that both boys and girls tend to rate the women teachers considerably lower than men teachers on W.

One explanation for this finding is that children are less accustomed to close, frequent rapport with men, and feel that men are less approachable. Objectively similar warmth behaviors of men and of women teachers may be so perceived that pupils over-estimate the strength of the W behaviors of the men teachers. Or perhaps in this sample the men teachers had, in general, warmer personalities than did the women.

What Implications can be Drawn from the High Intercorrelation of W and M?—In the construction of the scales measuring the teacher behavior variables of W and M, an attempt was made to distinguish clearly between them. As previously discussed, competent judges did distinguish between the two variables. Yet the intercorrelation of W and M (.66 for boys and .70 for girls) and the results of partial and multiple correlation analysis, indicate that W scores have little predictive power that is not also available through scores for M, but not vice versa. It could be suggested that pupils did not distinguish between M and W, due either to inadequate scale construction or to pupils' subjective halo effects in their perceptions of teachers. However, the fact that in responding to the W items, both boys and girls rated women teachers as significantly lower on W than men teachers, but did not so differentiate as regards M items, indicates that pupils did distinguish between warmth and intrinsic motivation behaviors. That is, the tendency for pupils to rate a given teacher somewhat similarly on the variables of warmth and of intrinsic motivation was due, in part at least, to the correspondence within the teacher of these two types of behaviors. This may be interpreted to mean that in the training of teachers, instructions in the utilization of intrinsic motivation are more likely to be successful if the student teacher already possesses the attributes of warmth, or can be helped to acquire them.

Summary and Conclusions

This research was structured within the general pattern of teacher competence studies, with the purpose of identifying some teacher behaviors that relate to desirable pupil learning. The selection of the teacher antecedent variables (warmth, demand, and intrinsic motivation) was based on rationale as to teacher behaviors that would logically contribute to the changing of pupil behaviors. The selection of the consequent, desirable pupil learning (interest in science), rested upon judgments as to what constitutes important, measurable school objectives.

The sample included 1045 ninth-grade boys and girls and their 38 general science teachers from 19 public schools in eastern Massachusetts.

The design of the research included control of four factors which might affect the criterion scores for classes: school subject, grade level, sex of pupil, and father's interest in science.

Pupils within a class agreed closely in their ratings on the variables of warmth and of intrinsic motivation, with reliabilities between .88 and .93 for the stability of within-class responses. Pupil agreement on the variable of demand was less high, with reliabilities between .78 and .80.

The Science Interest Inventory, a measure of voluntary science activities, provided the criterion scores. Its reliability, as determined by the split-half technique and corrected by the Spearman-Brown formula, is .97.

Analysis supported the conclusion that the classroom science teacher can expect pupils to agree in general as to how they perceive him; he can also expect some divergence of individual pupil impressions.

A test of the differences among the class (teacher-group) means of pupil ratings indicated in the teacher sample a diversity for each teacher behavior variable, significant at the .001 level.

A major difference was found between the means for the 584 boy and the 452 girl science interest scores, with boys reporting significantly more activities than girls (.001 level).

An hypothesis that the teacher variable of demand is non-linearly related to pupils' interest in science was rejected. No significant relationship was found for these two variables. It was suggested that the evidence from this research indicates that tensions over school work are not necessarily detrimental to such broad learning objectives as interest in the subject matter.

An hypothesis that the teacher variable of utilization of intrinsic motivation is positively related to pupils' interest in science was accepted. The within-classes correlation for boys was .32 and for girls was .42, with both significant at .001 level. Further analysis demonstrated that the 38 separate correlations for boy classes are equivalent, and the 34 separate correlations for girl classes are equivalent; any variations can be attributed to chance.

An hypothesis that the teacher variable of warmth is positively related to pupils' interest in science was accepted, with positive and highly significant (.001) within-classes correlations of .20 for the boys and .28 for the girls.

A multiple correlation analysis demonstrated that 25% of the variance in pupils' science interest scores can be accounted for by a knowledge of the variation in scores for the teacher's utilization of intrinsic motivation and for the father's interest in science.

It was hypothesized that the interaction of the teacher variables might weaken or strengthen one another's effect on the pupils' interest in science. Analysis, however, revealed no significant interaction effects.

The findings for both the boy and the girl samples were remarkably similar throughout the analysis—e. g., in the various correlation tests, and tests for differences in boy and girl perceptions of teacher behaviors. A difference between boy and girl samples was found in the strength of the relationship of father's interest in science to son's or daughter's interest.

Analysis of the intercorrelation of the two teacher variables of warmth and intrinsic motivation demonstrated a fairly strong relationship. Other analysis indicated that pupils did distinguish between these two variables; it is therefore reasonable to suggest that these two teacher characteristics tend to coexist within the teacher.

Analysis indicated that pupils perceive the women teachers in this sample as significantly lower (.001 level) on warmth than the men teachers. The number of teachers (38) was too small to permit any generalized inferences beyond this sample.

In conclusion, the findings of this research seem to warrant the generalization that the science interests of many pupils in this sample are independent of the low and moderate demands of the teacher, but are a function of the teacher's capacity to establish a relaxed interpersonal relationship with the pupil, and of the teacher's ability to utilize the educational principle of intrinsic motivation.

FOOTNOTES

* This article is derived from the author's doctorate thesis, completed at the Harvard Graduate School of Education and entitled *Pupils' Interest in Science as a Function of the Teacher Behavior Variables of Warmth, Demand, and Utilization of Intrinsic Motivation.*

1. There is growing evidence that pupil perceptions are in close agreement with projective-type measures of teacher traits. (Projective-type measures of personality variables are frequently considered more valid than other types of instruments.) Teachers' self-ratings show low relationships with either of these.

2. For a more detailed discussion of the bases for this and succeeding hypotheses, the reader is referred to the original dissertation.

3. At the ninth-grade level in Massachusetts public schools, the science course offered is almost exclusively general science.

4. This variable was of an exploratory nature and is not sufficiently germane to be discussed in this article.

5. These variations can be attributed to chance.

6. The questions previously asked about the relationships of M and I can be repeated for the relationships of W and I. Only the major features of the latter are necessary for present purposes.

7. While the within-classes correlation of F and I is moderately strong, the differences among $\overline{F}$ scores for classes are significant only at the .05 level. Adjusting for such differences does not reduce greatly the unaccounted-for variance of $\overline{I}$ scores for classes.

BIBLIOGRAPHY

1. Ackerman, W. I. "Teacher Competence and Pupil Change," *Harvard Educational Review*, XXIV (1954), pp. 273-89.
2. Amatora, Sister Mary, "Similarity in Teacher and Pupil Personality," *Journal of Psychology*, XXXVII (1954), pp. 45-50.
3. Anderson, H. H., and Brewer, J. E. *Studies of Teachers' Classroom Personalities, II.* (Stanford, California: Stanford University Press, 1946).
4. Barr, A. S. "Impressions, Trends, and Further Research," *Journal of Experimental Education*, XIV (1945), pp. 200-6.
5. Barr, A. S. "Teaching Competencies," in *Encyclopedia of Educational Research*, Edited by W. S. Monroe (New York: Macmillan Co., 1950), pp. 1446-54.
6. Barr, A. S., and Jones, E. R. "The Measurement and Prediction of Teacher Efficiency," *Review of Educational Research*, XXVIII (1958), pp. 256-61.
7. Bush, R. N. *The Teacher-Pupil Relationship* (New York: Prentice-Hall, 1954).
8. Callis, Robert. "The Efficiency of the Minnesota Teacher Attitude Inventory for Predicting Interpersonal Relations in the Classroom," *Journal of Applied Psychology*, XXXVII (1953), pp. 82-5.
9. Cogan, M. L. "Theory and Design of a Study of Teacher-Pupil Interaction," *Harvard Educational Review*, XXVI (1956), pp. 315-42.
10. Cogan, M. L. "The Behavior of Teachers and the Productive Behavior of Their Pupils: I. 'Perception' Analysis; II. 'Trait' Analysis," *Journal of Experimental Education*, XXVII (1958), pp. 89-124.

11. Cronbach, L. J. Essentials of Psychological Testing (New York: Harper and Brothers, 1949).
12. Curtis, F. D. A Digest of Investigations in the Teaching of Science, Vols. I, II, and III. (Philadelphia: P. Blakiston, 1926-39).
13. Darley, J. G., and Hagenah, Theda. Vocational Interest Measurement, Theory and Practice (Minneapolis: University of Minnesota Press, 1955).
14. Domas, S. J., and Tiedeman, D. V. "Teacher Competence: An Annotated Bibliography," Journal of Experimental Education, XIX (1950), pp. 101-218.
15. Ewens, W. P. "The Development and Standardization of a Preliminary Form of an Activity Experience Inventory: A Measure of Manifest Interest," Journal of Applied Psychology, XL (1956), pp. 169-74.
16. Goethals, G. W. "A Framework for Educational Research," Harvard Educational Review, XXVIII (1958), pp. 29-43.
17. Greene, E. B. Measurements of Human Behavior (New York: Odyssey Press, 1952).
18. Guba, E. G., and Getzels, J. W. "Personality and Teacher Effectiveness: A Problem in Theoretical Research," Journal of Educational Psychology, XLVI (1955), pp. 330-44.
19. Hebb, D. O. "Drives and the Conceptual Nervous System," Psychological Review, LXII (1955), pp. 243-54.
20. Jackson, R. W. B. Application of the Analysis of Variance and Covariance Method to Educational Problems, Bulletin No. 11 of the Department of Educational Research, University of Toronto, 1940.
21. Jahoda, M., and others. Research Methods in Social Relations with Special Reference to Prejudice, Parts 1 and 2 (New York: Dryden Press, 1951).
22. Janis, I. L., and Feshbach, Seymour. "Effects of Fear-Arousing Communications," Journal of Abnormal and Social Psychology, XLVIII (1953), pp. 78-92.
23. Lanzetta, J. T., and others. "The Effects of an 'Anxiety-Reducing' Medication on Group Behavior Under Threat," Journal of Abnormal and Social Psychology, LII (1956), pp. 103-8.
24. Lippitt, Ronald, and White, Ralph. "An Experimental Study of Leadership and Group Life," in Readings in Social Psychology, Edited by G. Swanson and others (New York: Henry Holt Co., 1952), pp. 340-55.
25. May, Rollo. The Meaning of Anxiety (New York: Ronald Press, 1950).
26. McCall, William. Measurement of Teacher Merit (Raleigh, N. C.: North Carolina State Superintendent of Public Instruction, 1952).
27. McNemar, Quinn. Psychological Statistics (New York: John Wiley and Sons, 1955).
28. Mitzel, H. E., and Gross, C. F. A Critical Review of the Development of Pupil Growth Criteria in Studies of Teacher Effectiveness (New York: Board of Higher Education of City of New York, 1956).
29. Mowrer, O. H. "A Stimulus-Response Analysis of Anxiety and its Role as a Reinforcing Agent," Psychological Review, XLVI (1939), pp. 553-65.
30. Murray, H. A., and others. Explorations in Personality (New York: Oxford University Press, 1938).
31. Reed, H. J. "An Investigation of the Relationship Between Teaching Effectiveness and the Teacher's Attitude of Acceptance," Journal of Experimental Education, XXI (1953), pp. 277-325.
32. Ryans, D. G. "Theory Development and the Study of Teacher Behavior," Journal of Educational Psychology, XLVII (1956), pp. 462-75.
33. Super, D. E. The Psychology of Careers (New York: Harper and Brothers, 1957).
34. Thelen, H. Experimental Research Toward a Theory of Instruction," Journal of Educational Research, XLV (1951), pp. 90-3.
35. Walker, H. M., and Lev, Joseph. Statistical Inference (New York: Henry Holt and Co., 1953).
36. Wert, J. E., and others. Statistical Methods in Educational and Psychological Research (New York: Appleton-Century-Crofts, (1954).
37. Withall, J. G. "Assessment of the Social-Emotional Climates Experienced by a Group of Seventh Graders as They Moved from Class to Class," Educational and Psychological Measurements, XII (1952), pp. 440-51.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 3, March 1961)

THE TRANSFER EFFECT OF A LEARNING PROGRAM IN SOCIAL CAUSALITY ON AN UNDERSTANDING OF PHYSICAL CAUSALITY*

ROLF E. MUUSS**
Goucher College
Towson, Maryland

VARIOUS STUDIES have demonstrated that participation in an experimental learning program designed to develop an understanding and appreciation of the nature of human behavior and the dynamic factors operative in social situations, does substantially increase knowledge of social causality (10, 13, 14, 16, 20). Social causality has been defined as an awareness of the dynamic complexity of human motivation, and as an understanding of the interacting nature of the forces that operate in human behavior and in social situations in general. It involves flexibility in thinking, and a willingness to see things from the viewpoint of others. A causally oriented person will suspend judgment until sufficient information is available. He realizes that his behavior has consequences, and that there are alternative ways of solving most social problems (12). He has a functional understanding of the concept of probability in respect to measurements and predictions. A more detailed theoretical discussion of the concept of social causality can be found in the literature (8). The staff of the Preventive Psychiatry Program at the State University of Iowa is in the process of developing a causally oriented learning program in an attempt to investigate to what extent social causality contributes to school adjustment and mental health (12).

The emphasis in the program is on social causality, that is, an understanding of the multiple causes of behavior, the effects of behavior, and the importance of thinking in probability terms when considering both causes and effects. The question arises whether Ss, who have participated in a learning program primarily designed to develop a more thorough understanding of social causality and human motivation, also develop a more thorough understanding of the factors that operate in the physical world and that help to explain natural phenomena.

In analogy to the definition of social causality, the concept of physical causality might be defined as:

1. An awareness of some of the forces operating in the physical world, and the common factors that cause an event or a phenomenon in the physical world,
2. An awareness that there are frequently multiple factors or a complexity of factors involved in a given event or phenomenon in the physical world,
3. An awareness of some of the common scientific procedures and methods of fact finding,
4. An awareness that knowledge of a given factor, force or dimension as represented by a measurement of it, is limited since any measure is only an approximation,
5. An awareness of the fact that if knowledge is based on approximations, predictions based on this knowledge are also only approximate,
6. An awareness of the fact that scientific knowledge grows and may change as new information is accumulated.

Hypothesis

Natural science teaching has not given much attention to the development of thinking in probability terms. In a study by Fitzgerald and Ojemann no correlation was found between scores on a probability concepts test and number of semesters of high-school natural science study. The authors conclude "... no relationship between an understanding of the dynamic conception of knowledge as measured by the test and number of semesters of science is indicated."(3) In view of this finding and since the experimental causal learning program attempts to develop a functional conception of probability, one might expect some transfer effect from the social causality learning program to probability thinking as applied to natural science problems. One might also expect some reinforcement in considering the complexity of factors and in the understanding of scientific procedures and methods of fact finding which natural science teaching has begun to include in its elementary school teaching content.

---

* The data for this study were collected while the author was on the staff of the Preventive Psychiatry Program, State University of Iowa. The preparation of this paper was supported by Goucher College and the Grant Foundation.

**The author is greatly indebted to Dr. R. H. Ojemann for valuable advice in the preparation of the manuscript, and to Dr. B. Snider for assistance in the statistical analysis and IBM processing.

Thus, the major hypotheses of this study are:

1. Subjects who have participated in an experimental learning program designed to develop an understanding and appreciation of the forces that operate in human behavior will be more capable of solving problems involving an understanding of "causality" and "probability" than control subjects who have not had special training in these concepts.

2. It is further hypothesized that there is a transfer effect in that an individual who has learned to solve social problems dynamically will also approach phenomena and events in the physical world with a causal attitude, asking such questions as: How did this event develop? What are the factors involved? How accurate are our measurements? And, what are the probable consequences?

The data gathered also make possible the testing of some additional hypotheses:

3. Studies have shown (15, 19) that the environment in which children grow up contains several influences which teach a non-causal orientation toward behavior, while the predominant attitude toward natural science is a causal one. It is thus hypothesized that measurements of intelligence correlate less with measurements of social causality than with measurements of physical science knowledge.

4. Since intelligence facilitates problem-solving, the measurements of causality will be correlated to intelligence. Brighter children usually have a better knowledge of the antecedent factors that bring about an event; they have a better understanding of the probabilistic nature of knowledge and the changes in knowledge. It is hypothesized that the correlation between intelligence and measurements of causality will be greater for control Ss who have not had a special training than for experimental Ss. Sensitivity to the factors that bring about an event, even though influenced by intelligence, is greatly dependent on other factors, especially those learning experiences which are provided in the experimental causal learning program.

5. Since it is common knowledge that increase in age, signifying accumulation of experience, is an important variable in concept formation (21), it is hypothesized that there will be a developmental increase in scores on the measures of causality from fifth to sixth grade. The results would be confounded by at least two factors, age and special teaching methods for the experimental Ss. Therefore, the direct developmental increase ought to be demonstrated on the control Ss only. In our culture a causal approach is taught toward the natural science phenomena. On the other hand, there is evidence (15) that magazines and newspapers fail to make an analytic or dynamic approach to the problems of development and neglect to describe effective methods of dealing with differential causes of behavior and variability in the rate of development. Furthermore, it has been shown (19) that the social studies readers which are used in the elementary school place emphasis on the external forms of behavior rather than the dynamics underlying it. Consequently, one might further hypothesize that for the control Ss the developmental increase in physical causality will be greater than in social causality.

The rationale for these hypotheses is based on the assumption that there is no basic dichotomy between social and natural phenomena. The causal question "Why does a given phenomenon happen?" is involved in both areas. Furthermore, the probabilistic nature of knowledge, the awareness that knowledge changes and expands, is recognized in the natural as well as in the social sciences. The inaccuracy of measurements and of predictions is equally applicable to both social and natural phenomena, even though the error term is likely to be greater in measurements used by the social scientist. The basic principles involved in an understanding of cause-effect relationships are equally applicable in the physical and in the social world. Consequently, there is justification in hypothesizing that an experimental teaching program which is designed to teach a basic understanding of principles of social causality will have a transfer effect on phenomena generally considered as part of the physical world.

The Test Instruments

The following test instruments were used to evaluate the Ss' knowledge and understanding in the area of social and physical causality:

1. The Elementary Causal Test (ECT) which consists of 30 items of the true-false type. The test measures the child's awareness of the dynamic, complex, variable nature of human behavior (16). The test has a Kuder-Richardson reliability of .63 and a correlation with intelligence of -.29 (N = 245) for sixth and -.37 (N = 158) for fifth-grade Ss.

2. The Problem Situations Test (PST) which

consists of 22 multiple choice items describing interaction between peers, siblings, parents, and teachers. The test measures the child's willingness to be immediately punitive in a hypothetical situation where no retaliation is anticipated. The situations involve aggressive feelings, moral transgressions and personal problems (7, 9, 16). The test has a Kuder-Richardson reliability of .77 and a correlation with intelligence of -.12 (N = 245) for sixth-grade, and -.12 (N = 158) for fifth-grade Ss.

For both the ECT and the PST the lower score is the more causal score.

3. In the following year when the study was repeated, with some modifications to be described later, the Problem Series III Test (PS III) was utilized as a measure of social causality. This change in the test instrument appeared to be necessary since the experimental Ss approached the ceiling of the PST and the ECT. The Problem Series III test consists of 62 true-false items involving various aspects of the concept of causality in social situations. Answers can be classified into four levels or degrees of causal understanding:

1) No concern or awareness as to behavior dynamics
2) Pseudo-causal approaches
    a. Using mystical or magical explanations or
    b. Using over-generalized observation or similar untested assumptions (stereotyping, rationalizing)
3) Recognizing one, or a limited number of observable factors
4) Recognizing a complex causation, searching for the most probable hypothesis (17).

The problem Series III test has a Kuder-Richardson reliability of .73. Its correlation with intelligence is -.25 (N = 113) for fifth-grade Ss. The test-retest reliability with one school year intervening (fall 1958 to spring 1959) is .49 (N = 113) for fifth-grade Ss; broken down for experimental and control groups the respective test-retest reliabilities are .59 (N = 51) and .56 (N = 62).

4. The Physical Causal Test (PCT) consists of 60 multiple choice items and was designed by the author to measure a subject's understanding of physical causality as defined above. The test has a Kuder-Richardson reliability of .80 and a correlation with intelligence of .58 (N = 245). The test is administered in two parts and subscores can be utilized from each of the two parts. Physical Causal Test Part I (PCT I) consists of 38 items and mainly measures knowledge of causal explanations of common phenomena and events in the physical world. The test involves predominantly the first two points of our definition on page 231. The Kuder-Richardson reliability of this subscore is .76. The Physical Causal Test Part II (PCT II) consists of the remaining 22 items. Many of these items were adopted by the author from Clark's doctoral dissertation (1) for use with fifth- and sixth-grade Ss. The test in line with the definition on page 231 involves concepts three to six: scientific methods, the approximate nature of measurements, the approximate nature of predictions and an awareness of the changing nature of knowledge. While Part I is predominantly concerned with knowledge of phenomena and events in the physical world, Part II emphasizes the probalistic nature of scientific knowledge. Part II, due to its shortness, has a Kuder-Richardson reliability of .66.

5. In the succeeding year when the study was repeated this Physical Causal Test was revised. The total test of 60 items was divided into two forms, Form A and Form B, of 30 items each. The content of the items, the degree of difficulty and the discriminatory power of the items used in Form A and Form B were as similar as feasible, so that two equivalent forms of the test were available. The test-retest reliability of the two test forms over a two-week interval is .74 (N = 72). The correlation with intelligence is .45 (N = 116) for Form A and .44 (N = 116) for Form B.

Since the Physical Causal Test has not been described in the literature, a sample item will be given for each of the six concepts of the definition of physical causality.

1. An awareness of factors that cause a phenomenon.
   What could you do to make the water in the air form in drops on the outside of a glass?
   A. Set a glass of ice water in a very warm room.
   B. Set a glass of ice water in the refrigerator.
   C. Set a glass of warm water in a very warm room.
   D. Set an empty glass in a very warm room.
   E. Set a glass of hot water in a very cold room.

2. An awareness of the multiplicity of factors that cause a phenomenon.
   Not all people require the same number of food calories or food energies each day because:
   A. They do not like to eat the same thing.
   B. No two people do the same kind of work.
   C. Some days a person is hungrier than on other days.
   D. People burn up different amounts of calories depending whether they are a boy or girl and where they are.
   E. People burn up a different amount of calories depending on what they are doing and how much they weigh.

3. An awareness of common scientific procedures and methods of fact finding.

A party of explorers made a trip to northern Alaska to explore that country. Though they lived there for a month in freezing and snowy weather without enough warm clothing, none of the men had any colds. Then they received a shipment of warm clothing. A few days after opening the boxes of clothing, more than half the men came down with bad colds. If you were a scientist interested in the causes of colds, what conclusion would you draw from this situation?

A. It proves that colds are caused by a germ that can be carried in clothing.
B. It shows that it might be a good idea to make some experiments to see if cold germs can be carried in clothing.
C. It proves that cold, dampness and not enough clothing will eventually, if not right away, cause colds.
D. It shows that bad weather and not enough clothing are not causes of colds.
E. It does not have any meaning for a scientist because it was not a scientifically made experiment.

4. An awareness that measures are only approximations.

Two men were measuring the strength of the grips of the pupils in a sixth-grade class. Each pupil had three trials on Monday and another three on Tuesday. The conditions were about the same both days. For some of the children the average on Tuesday was as much as two pounds higher than the day before and for several others the average weight was as much as two pounds lower. With which statement do you agree?

A. Examiners can always read the scale exactly right but children's strength of grip isn't always the same.
B. Some differences should be expected from one day to the next.
C. They didn't get good measurements of those who weren't the same both days.
D. They got exact measurements for those who were the same both days.
E. The children had the same strength both days but the examiners read the scale carelessly.

5. An awareness that predictions are only approximations.

A well-known manufacturer of car tires produced a new tire. After many tests, he concluded the tire could be driven for at least 40,000 miles before it would be worn out. He put the tire on the market with a guarantee that the tire would last for that mileage, unless there were accidents and unusual driving conditions. Jane's father traded in all his old tires on a purchase of these new tires. What statement about his tires would you say would be most likely true after they had been driven 39,000 miles?

A. Probably all of these tires will be useable at 39,000 miles if there were no accidents or unusual driving conditions, but it is certain that at least one of them will still be useable.
B. If there were no accidents or unusual driving conditions, all of these tires will still be useable at 39,000 miles.
C. If there were no accidents and unusual driving conditions, it is likely that these tires will still be useable at 39,000 miles, but it is also possible that all of them will be worn out.
D. If there were no accidents or unusual driving conditions, it is likely that some of the tires will still be useable at 39,000 miles, but it is certain that at least one of them will be worn out.
E. Probably all of these tires will be worn out at 39,000 miles because tires very seldom last that long. The guarantee wouldn't mean much because it would be hard to prove that there had been no accidents or unusual driving conditions.

6. An awareness of the change and increase in knowledge.

A student who wanted to become a farmer was studying in school how to keep a farm so it will grow good crops. He studied his book thoroughly and took good notes on the lectures so that he could always use this knowledge in keeping the soil of his farm fertile. The textbook was the best the teacher could get at the time and he presented information accurately. After the student finished school, four years passed before he became a manager of a farm. When he did become a manager, he applied to his farm all the suggestions given in the course in soil fertility and followed them very carefully. He has been doing this for the last five years. With which statement do you agree?

A. It was a good idea to follow very carefully what he had learned in the class in soil fertility.
B. His way could be wrong because it would be old and out-of-date.
C. He should check up and see about new information about soil fertility that may have come out.
D. It would have been better to get a brand new book on soil fertility and have followed it exactly because new ways are best.
E. It would be a good idea to find out about any changes in knowledge because in nine years there are sure to be improvements in every field.

Procedures

The study involves two independently obtained sets of data, the first involved 413 fifth- and sixth-grade Ss during the 1957-58 school year; the second involved 113 fifth-grade Ss in the 1958-59 school year. The procedures for both aspects of the study will be described separately.

In the spring of 1958, the following tests were administered to 158 fifth- and 255 sixth-grade Ss in a midwestern community of 80,000:

1. The Elementary Causal Test (N = 403)
2. The Problem Situations Test (N = 403)
3. The Physical Causal Test (N = 413)

Of these 413 Ss, 208 had participated in an experimental learning program designed to develop a causal orientation toward human behavior and social problems. The remaining 205 Ss came from regular classrooms of the same school system. An attempt was made to match the experimental classes with the control classes on such variables as the teacher's professional training, the teacher's years of professional experience, the teacher's age, and the socio-economic background of the students.

The causal learning program conducted under the auspices of the staff of the Preventive Psychiatry Program, State University of Iowa, has been in operation for several years. Some aspects of the teacher training program and the content material utilized in this learning program have been described elsewhere in the literature (16).

Since the experimental and the control classes differed in their respective mean IQ scores (sixth-grade experimental 111.1 and control 106.1; fifth-grade experimental 112.0 and control 106.4) and since the Physical Causal Test has a moderately high and significant correlation with IQ (Tables III and IV), a simple analysis of covariance design was utilized in order to control statistically the concomitant variable of intelligence. As a measure of intelligence the Ss' latest Otis IQ scores were obtained from the school's cumulative folders.

The three measures of causality were administered at the end of the 1957-58 school year. It is assumed that differences which arose between experimental and control classes can be attributed to the special learning program to which the experimental Ss had been exposed. In order to test whether this assumption is justified, the experiment was repeated with 113 fifth-grade Ss in the 1958-59 school year. In this second part of the study the following tests were administered:

1. The Problem Series III, administered in the fall of 1958 and re-administered in the spring of 1959.
2. The Physical Causal Test Form A in the fall of 1958.
3. The Physical Causal Test Form B in the spring of 1959.

The scores obtained during the administration in the fall of 1958 will be referred to as the "pre-test scores"; the scores obtained in the spring of 1959 as the "post-test scores", and the differences between the pre- and post-tests as the "growth scores." All test scores plus the Otis IQ from the school's cumulative folders were available for 51 experimental and 62 control Ss. None of the experimental Ss had been in the experimental program before. Since there were no significant differences between IQ scores (Table VI) the pre-test scores and growth scores were analyzed by way of t tests.

Results and Discussion of Findings

The data pertaining to the first and the second hypotheses comparing experimental and control Ss on the criterion variables are reported in Table I for the sixth-grade Ss and in Table II for the fifth-grade Ss. Table I reports the analysis of covariance for 245 sixth-grade Ss for the ECT and the PST and also for 255 sixth-grade Ss for the PCT. Table II presents the corresponding data for 158 fifth-grade Ss. Tests for homogeneity of variance for all six sets of data in Table I and Table II as well as the data reported in Table VIII are evidence that we are justified in the necessary assumption of homogeneity of variance.

The data obtained from the two measures of social causality basically support the first hypothesis. For the Elementary Causal Test the F ratios are 19.89 for the sixth-grade Ss (Table I) and 32.73 for fifth-grade Ss (Table II). These differences are highly significant and give substantial support to the hypothesis that on the Elementary Causal Test the experimental Ss respond significantly more causally than the control Ss.

The data from the Problem Situations Test are not quite as convincing, the F ratio for sixth-grade Ss is 1.94, not significant (Table I), and 22.78, significant at the .001 level for fifth-grade Ss (Table II). It can be stated that on the Problem Situations Test fifth-grade experimental Ss respond more causally and less punitively than do their control Ss. The data does not warrant a similar conclusion for the sixth-grade Ss. However, inspection of individual scores indicates that many experimental Ss approach and reach the ceiling, that is, the lowest possible score of this test. On an average, experimental Ss answer non-causally only four out of twenty-two items which means that a number of Ss answer only zero, one or two items non-causally. They have little or no opportunity to improve further and to appropriately show their knowledge of causality. Consequently their knowledge and understanding of causality is

TABLE I

SUMMARY TABLE OF ANALYSIS OF COVARIANCE OF THE CAUSAL TESTS FOR SIXTH-GRADE EXPERIMENTAL AND CONTROL SUBJECTS (1957-58 DATA)

Sources	df	SSX	SPXY	SSY	SS'Y	df	ms'Y	F
Elementary Causal Test								
Methods (A)	1	1337. 4	846. 8	536. 1	361. 18	1	361. 18	19. 89**
Within (W)	243	25365. 2	2534. 8	4646. 7	4393. 39	242	18. 15	
Total (T)	244	26702. 6	3381. 6	5182. 8	4754. 56	243		
Problem Situations Test								
Methods (A)	1	1337. 4	263. 3	41. 8	25. 79	1	25. 79	1. 94
Within (W)	243	25365. 2	867. 8	3241. 0	3211. 26	242	13. 27	
Total (T)	244	26702. 6	1104. 2	3282. 7	3237. 05	243		
Physical Causal Test								
Methods (A)	1	1580. 0	1192. 1	899. 5	168. 88	1	168. 88	4. 13*
Within (W)	253	26052. 1	10884. 9	14848. 3	10300. 42	252	40. 87	
Total (T)	254	27632. 1	12077. 0	15747. 8	10469. 29	253		

* Significant at the . 05 level
**Significant at the . 001 level

TABLE II

SUMMARY TABLE OF ANALYSIS OF COVARIANCE OF THE CAUSAL TESTS FOR FIFTH-GRADE EXPERIMENTAL AND CONTROL SUBJECTS (1957-58 DATA)

Sources	df	SSX	SPXY	SSY	SS'Y	df	ms'Y	F
Elementary Causal Test								
Methods (A)	1	1252.0	938.0	702.8	491.24	1	491.24	32.73**
Within (W)	156	21221.9	2220.1	2558.6	2326.37	155	15.01	
Total (T)	157	22473.9	3158.1	3261.4	2817.61	156		
Problem Situations Test								
Methods (A)	1	1252.0	580.0	268.6	241.66	1	241.66	22.78**
Within (W)	156	21221.9	235.7	1646.9	1644.31	155	10.61	
Total (T)	157	22473.9	815.6	1915.6	1885.98	156		
Physical Causal Test								
Methods (A)	1	1252.0	1011.5	817.2	249.83	1	249.83	6.91*
Within (W)	156	21221.9	7389.1	8178.5	5605.77	155	36.17	
Total (T)	157	22473.9	8400.6	8995.7	5855.60	156		

* Significant at the .01 level
**Significant at the .001 level

not measured adequately by this test.

The tendency for experimental Ss to reach the ceiling of the Problem Situations Test after participating in a causal learning program has been found in previous studies (16). This fact made it necessary to develop a new test instrument for the measurement of social causality. The Problem Series III was developed as a result of this finding and used in the follow-up study in 1958-59. If one considers this particular weakness of the PST which becomes most obvious for sixth-grade Ss, the data substantially support the hypothesis that experimental and control Ss do differ in their responses to these two measures of social causality. The experimental Ss respond more causally and less punitively. Since intelligence is controlled by the method of covariance it is assumed, in agreement with other studies, that the obtained differences can be attributed to the learning program.

The data for the second hypothesis are also reported in Table I and Table II. As a measure of transfer, the Physical Causal Test, which is relatively unrelated to the learning program in its content but is related in the method of problem solving, was utilized. The F ratios are 4.13, significant at the .05 level for sixth-grade Ss, and 6.91, significant at the .01 level for fifth-grade Ss. These data demonstrate convincingly that the experimental Ss had a better understanding of factors that bring about an event in the physical world and were more aware of the probabilistic nature of knowledge than their control Ss. It is inferred that this difference is due to a transfer effect from the experimental learning program. Judd's theory of transfer of training (5) states that learning transfers through the understanding of underlying principles and the formation of generalizations. One might argue that some of the basic principles involved in the solving of these test problems are analogous to the concepts taught in the experimental learning program but applied in the PCT test to a different content.

The coefficients of correlation pertaining to the third hypothesis can be found in Tables III, IV, and VII. As a word of explanation, the reader is reminded that the measures of social causality—the ECT, the PST, and PS III—are scored in such a way that the *lower* score is the more causal score, while on the PCT the *higher* score is the more causal score. This fact results in negative correlations between social and physical causal scores and between social causal scores and IQ. Consequently, it is the magnitude of the correlation and not its direction that matters in the interpretation of findings.

For sixth-grade Ss (N = 245) the PCT/IQ correlation is .58, on the other hand, the PST/IQ correlation is only -.12, and the ECT/IQ correlation is -.29. Disregarding the negative sign, for reasons explained above, Fisher's *z* test of differences between *r*'s indicates in support of the third hypothesis that the PCT/IQ correlation is significantly greater than both the ECT/IQ and the PST/IQ correlations.

The correlations for fifth-grade Ss (N = 158) are quite similar in magnitude to those for sixth-grade Ss—PCT/IQ is .59, PST/IQ is -.12, and ECT/IQ is -.37. Fisher's *z* test of differences between *r*'s again evidences that the PCT/IQ correlation is significantly greater than both the ECT/IQ and the PST/IQ correlations. The hypothesis that measures of social causality are not as highly correlated to intelligence as those of physical causality is substantiated by the 1957-58 data.

In taking a preview of the 1958-59 data in respect to this hypothesis (Table VII), we observe that the PCT Form A/IQ correlation obtained in the fall of 1958 is .45 (N = 116) and the Form B/IQ correlation obtained in the spring of 1959 is .44 (N = 116). The PS III/IQ correlation is -.36 for the fall of 1958 and -.25 for the spring of 1959 administrations. The differences between the two spring correlations approach the .10 level of significance. These data do not repudiate the 1957-58 findings.

It appears that the solving of natural science problems is more enhanced by intellectual ability than is the solving of problems involving social conflicts, motivation of behavior, and a willingness to understand others. There are two possible explanations for this finding: a) The understanding of physical science phenomena is taught as an intellectual subject and items similar to those of the PCT are not uncommon in intelligence tests. The ability to understand the behavior of others is often considered to be related to one's own emotional make-up and one's personality rather than to one's intellect. b) An analytic and causal approach to natural phenomena is part of our intellectual culture and expectations, while many segments of our culture are more concerned with the overt forms of behavior rather than its underlying dynamics and motives.

It was assumed in hypothesis four that the measures of causality show a positive correlation with intelligence. This assumption is justified since the respective correlations are significant for the ECT and the PCT but not for the PST (Tables III and IV). For the 1958-59 data, all respective correlations are significant.

The second part of the fourth hypothesis relates to differences between experimental and control Ss in their respective correlations between IQ and the criterion variables. In comparing the IQ/criterion variables correlations between the experimental and the control Ss, one observes, in agreement with the hypothesis, that five out of six (Tables III and IV) are in the predicted direction. The PST/IQ correlation for fifth-grade Ss constitutes the only exception. However, only in one of five instances, namely the PST/IQ correlation for

TABLE III

INTERCORRELATIONS OF THE CRITERION VARIABLES AND IQ FOR SIXTH-GRADE SUBJECTS (1957-58 DATA)

		Elementary Causal Test	Problem Situations Test	Physical Causal Test
Intelligence Quotient				
Sixth-Grade Ss	(N = 245)	-.29**	-.12	.58**
Experimental Ss	(N = 127)	-.17	.02	.53**
Control Ss	(N = 118)	-.29**	-.21*	.58**
Elementary Causal Test				
Sixth-Grade Ss	(N = 245)	....	.51**	-.48**
Experimental Ss	(N = 127)	....	.56**	-.34**
Control Ss	(N = 118)	....	.45**	-.53**
Problem Situations Test				
Sixth-Grade Ss	(N = 245)	....	....	-.23**
Experimental Ss	(N = 127)	....	....	-.14
Control Ss	(N = 118)	....	....	-.29**

* Significant at the .05 level
**Significant at the .01 level

TABLE IV

INTERCORRELATIONS OF THE CRITERION VARIABLES AND IQ FOR FIFTH-GRADE SUBJECTS (1957-58 DATA)

		Elementary Causal Test	Problem Situations Test	Physical Causal Test
Intelligence Quotient				
Fifth-Grade Ss	(N = 158)	-.37**	-.12	.59**
Experimental Ss	(N = 75)	-.23*	-.15	.53**
Control Ss	(N = 83)	-.35**	.06	.59**
Elementary Causal Test				
Fifth-Grade Ss	(N = 158)	....	.52**	-.55**
Experimental Ss	(N = 75)	....	.58**	-.35**
Control Ss	(N = 83)	....	.34**	-.57**
Problem Situations Test				
Fifth-Grade Ss	(N = 158)	....	....	-.14
Experimental Ss	(N = 75)	....	....	-.08
Control Ss	(N = 83)	....	....	.01

* Significant at the .05 level
**Significant at the .01 level

sixth-grade Ss, does the difference approach the .05 level of significance. Anticipating the 1958-59 data (Table VII) one observes that in the fall administration the PCT/IQ correlations of the experimental and those of the control groups are not significantly different. In the spring administration, after the experimental Ss had been in the program for nine months, there is a significant difference; control Ss show a significantly higher PCT/IQ correlation than experimental Ss, thus supporting the hypothesis. The respective data for the PS III demonstrate a significant difference between experimental (r = -. 13) and control (r = -.52) Ss for the fall, while the difference for the spring administration (r = -. 12 experimental; r = -.40 control) approaches the . 10 level of significance.

Summarizing the discussion relating to the fourth hypothesis, differences between experimental and control Ss are in the predicted direction; however, they are not consistently significant, and, therefore, give only tentative support for the hypothesis. Thus while it is generally true that the more intelligent Ss have a better understanding of cause-effect relationships, the probabilistic nature of knowledge, etc., this relationship, as was hypothesized, appears to be more pronounced for Ss who have not participated in an experimental learning program. The data, especially from the 1958-59 study, suggest that the experimental learning program reduces the relationship between intelligence and causal understanding. This finding is in agreement with the data reported in Tables I and II, namely that the causal learning program increases a S's awareness of what has been defined as 'causality' regardless of his intelligence. There even appears to be a negative though non-significant relationship between intelligence and the growth score on the Physical Causal Test for the experimental Ss (r = -.15, N = 51).

The fifth hypothesis, supported by various developmental studies on causality (2, 4, 6, 18, 21), states that children as they grow older become more capable of understanding the physical and the social world in which they live due to their accumulated knowledge and experience. In order to test this hypothesis, t tests were computed on the criterion variables between the fifth- and sixth-grade control Ss. In this instance, the t test suffices since there are no statistically significant differences between the fifth- and sixth-grade Ss in their mean Otis intelligence scores (Table V). No further importance will be attached to this since there would be—for obvious reasons—differences if the mental age had been used instead of the IQ.

Table V reports the data for 201 control Ss. In support of the hypothesis there are significant differences on the Physical Causal Test. An understanding of causality as measured by the PCT is enhanced by an increase in age and experience. There also is a significant difference on the Problem Situations Test, while the difference on the Elementary Causal Test approaches the . 05 level of significance. Consequently, the data from the three tests of causality support the hypothesis that sixth-grade Ss have a better understanding of causality than fifth-grade Ss.

The data pertaining to the second part of the hypothesis are not conclusive, however, it is suggested by the magnitude of the t ratios that developmental increase in causality is greater on the PCT than on the ECT for which the t ratio is not quite significant.

Modification of the Study

The question arises as to what extent the experimental and control Ss differed originally, and whether the obtained differences—as was assumed—are due to the causal learning program and not to some other factors. Since, as was shown, intelligence is an influential factor, its effect was eliminated by way of the statistical method of covariance.

In order to further support our basic hypothesis and to control the effect of the experimental learning program more effectively, the study with some modification was repeated the following school year with fifth-grade Ss. Since there are no significant differences in intelligence between the experimental and the control Ss, the data in Table VI are reported by way of t tests between group means. In the 1958-59 administration the Physical Causal Test Form A was administered in the fall of 1958 and the Form B in the spring of 1959. As a measure of social causality the Problem Series III test was administered in the fall of 1958 and again in the spring of 1959. This makes it possible to compare the experimental and the control Ss at the beginning of the school year and their growth of causality during the school year. Table VI supplies the 1958-59 data for the 113 fifth-grade Ss. At the beginning of the fall 1958 school year, there were no significant differences on the pre-test for the PS III and the PCT; neither were there significant differences in intelligence. Consequently, there is rather conclusive evidence for the assumption that the experimental Ss did not differ from the control Ss at the beginning of the 1958-59 school year on the criterion variables under investigation.

The effect of the experimental learning program was measured in terms of growth scores on the PCT and the PS III.

Even though the growth score was utilized and not the terminal score, as in the 1957-58 analysis, findings are identical. In agreement with our first hypothesis, the data indicate that even though the two groups do not differ significantly on the pretest of the Problem Series III, the experimental Ss show significantly more growth, t ratio 5. 75

TABLE V

COMPARISON OF THE MEAN SCORES OF FIFTH- AND SIXTH-GRADE CONTROL SUBJECTS ON THE CRITERION VARIABLES (1957-58 DATA)

	Sixth-Grade Ss			Fifth Grade Ss			
	N	Mean	SD	N	Mean	SD	t
Intelligence Quotient	118	106. 13	10. 32	83	106. 35	11. 61	. 14
Elementary Causal Test	118	8. 22	4. 91	83	9. 53	4. 63	1. 92*
Problem Situations Test	118	4. 88	3. 69	83	6. 14	3. 32	2. 52**
Physical Causal Test	118	29. 49	7. 38	83	26. 45	7. 39	2. 86***

* Significant at the . 10 level
** Significant at the . 05 level
***Significant at the . 01 level

TABLE VI

COMPARISON OF THE MEAN SCORES OF THE EXPERIMENTAL AND CONTROL FIFTH-GRADE SUBJECTS ON IQ AND THE CRITERION VARIABLES (1958-59 DATA)

	Experimental Ss			Control Ss			
	N	Mean	SD	N	Mean	SD	t
Intelligence Quotient (Otis)	51	108. 55	7. 44	62	109. 37	11. 34	. 46
Pre-Test Fall 1958 Problem Series III	51	21. 86	5. 31	62	20. 65	4. 84	1. 25
Growth Scores Problem Series III Fall-Spring Score	51	7. 57	4. 86	62	2. 40	4. 54	5. 75**
Pre-Test Fall 1958 Physical Causal Test Form A	52	12. 29	3. 75	65	13. 28	3. 90	1. 38
Growth Scores Physical Causal Test Form B - Form A	52	2. 88	3. 42	65	1. 42	3. 65	2. 22*

* Significant at the . 05 level
**Significant at the . 001 level

(significant at the .001 level), on this measure of social causality than the control Ss. Furthermore, in support of the second hypothesis one observes that even though both groups do not differ significantly on the Physical Causal Test at the time of the pre-test, the experimental Ss show a significantly greater mean gain during the two test administrations than do the control Ss, thus again supporting the hypothesis concerning a transfer effect from the experimental learning program to a measure of physical causality.

In conclusion, the data from the 1958-59 study give additional support to the hypothesis that experimental Ss grow more rapidly in their understanding of causality than control Ss. Since it was shown that the two groups did not differ significantly on the criterion variables at the beginning of the experiment, it can be assumed that the growth in causality is a product of the experimental causal learning program. Furthermore, there is justification in concluding that the experimental learning program with its emphasis on dynamics of behavior and human motivation does contribute to a transfer effect which enhances the experimental Ss' performance in solving problems involving an understanding of common scientific phenomena, an awareness of the multiplicity of factors that may be involved in bringing about an event, an awareness that knowledge may change, an understanding of error terms in measurement and prediction, and an appreciation of the probabilistic nature of scientific knowledge.

The findings of a previous study (11) suggest a further problem. Is the transfer effect as measured by the Physical Causal Test greater for the kind of items that are presented in PCT Part I involving knowledge of common phenomena and events in the physical world or is it greater for the items in PCT Part II which emphasize the probabilistic nature of knowledge? Furthermore, if there is a differential transfer effect, is it the same for fifth- and for sixth-grade Ss?

In order to investigate this question the scores of the Physical Causal Test reported in Table I and Table II were broken down into their subscores, PCT Part I and PCT Part II. In Table VIII these subscores are analyzed by a covariance design in order to eliminate the influence of IQ and are reported separately for fifth- and sixth-grade Ss. The data for sixth-grade Ss, in agreement with previous findings (11), demonstrate that there is no significant difference between experimental and control Ss on the PCT Part I. On the other hand, the $\underline{F}$ ratio for the PCT Part II is 12.36, significant at the .001 level. Experimental sixth-grade Ss have a significantly better understanding of the probabilistic nature of knowledge than do control Ss, but do not differ in the knowledge of common phenomena and events in the physical world as measured by these tests.

For fifth-grade Ss the findings are reversed. The PCT Part I scores yield an $\underline{F}$ ratio of 11.68, significant at the .001 level, indicating that even though there is no systematic difference between sixth-grade experimental and control Ss, such a difference exists for fifth-grade Ss. On the other hand, while the difference was highly significant for sixth-grade Ss on the PCT Part II, there is no significant difference on this subscore for fifth-grade Ss.

It appears that there is a treatment and grade level interaction effect, which only becomes obvious when the total Physical Causal Test score is broken down into its two subscores. Fifth-grade experimental Ss are better able to solve the kind of problems that make up the PCT Part I, while sixth-grade experimental Ss solve the problems of the PCT Part II more effectively than their respective control Ss.

The differential findings on the PCT Part II can be explained fairly easily by the concepts involved. The items making up the PCT Part II are relatively difficult and complex. Many of these items were originally developed by Clark (1) for use with junior high-school Ss and were adopted by the author for use with fifth- and sixth-grade Ss. While the adaptation for sixth-grade Ss appears to have resulted in an appropriate degree of difficulty, the items do seem to be fairly difficult for both experimental and control fifth-grade Ss, and the experimental learning effect does not show on these items. Furthermore, the particular content of the experimental learning program may contribute to the findings on the PCT Part II, for there is a greater emphasis in the sixth-grade material on concepts involving probability, change of knowledge, and the inaccuracy of measurement and predictions.

The findings pertaining to the PCT Part I are more difficult to explain. It appears that for the fifth-grade experimental Ss there is a real learning and/or transfer effect which increases the PCT Part I subscore and which is assumed to be due to the experimental learning program. Why this is not the case for sixth-grade Ss is not obvious. Perhaps the PCT Part I items are too direct, many items ask for relatively simple explanations and at times only for one or two factors that may have caused an event. The sixth-grade experimental Ss might have reached a level of sophistication in their causal thinking, particularly in respect to the multiplicity of causes and the interacting nature of factors that bring about an event, so that they were confused by the rather direct explanations asked for in the PCT Part I. Also the sixth-grade control teacher, with the post-Sputnik emphasis on natural science teaching in the curriculum, may put greater emphasis on science teaching and since the concepts covered in PCT Part I would be taught in elementary school in a conventional physical science class, such science teaching would indeed help a student in solving the

TABLE VII

INTERCORRELATIONS OF THE CRITERION VARIABLES AND IQ FOR FIFTH-GRADE SUBJECTS (1958-59 DATA)

		Pre-Test Fall 1958 PS III	Post-Test Spring 1959 PS III	Pre-Test Fall 1958 PCT Form A	Post-Test Spring 1959 PCT Form B
Intelligence Quotient					
Fifth-Grade Ss	(N = 113)	-.36**	-.25**	.44**	.43**
Experimental Ss	(N = 51)	-.13	-.12	.30	.16
Control Ss	(N = 62)	-.52**	-.40**	.52**	.58**
Pre-Test Fall 1958 PS III					
Fifth-Grade Ss	(N = 113)	....	.49**	-.38**	-.39**
Experimental Ss	(N = 51)	....	.59**	-.30*	-.32*
Control Ss	(N = 62)	....	.56**	-.43**	-.46**
Post-Test Spring 1959 PS III					
Fifth-Grade Ss	(N = 113)	....	....	-.31**	-.39**
Experimental Ss	(N = 51)	....	....	-.40**	-.37**
Control Ss	(N = 62)	....	....	-.40**	-.44**
Pre-Test Fall 1958 PCT Form A					
Fifth-Grade Ss	(N = 113)	....	....	....	.57**
Experimental Ss	(N = 51)	....	....	....	.57**
Control Ss	(N = 62)	....	....	....	.58**

* Significant at the .05 level
**Significant at the .01 level

TABLE VIII

SUMMARY TABLE OF ANALYSIS OF COVARIANCE OF THE SUBSCORES FOR THE PHYSICAL CAUSAL TEST PART I AND PART II FOR FIFTH- AND SIXTH-GRADE EXPERIMENTAL AND CONTROL SUBJECTS

Sources	df	SSX	SPXY	SSY	SS'Y	df	ms'Y	F
Physical Causal Test Part I (Sixth-grade Ss)								
Methods (A)	1	1580.0	509.6	164.4	2.06	1	2.06	.10
Within (W)	253	26052.1	7439.8	7553.0	5428.36	252	21.54	
Total (T)	254	27632.1	7949.5	7717.4	5430.42	253		
Physical Causal Test Part II (Sixth-grade Ss)								
Methods (A)	1	1580.0	682.5	294.8	133.80	1	133.80	12.36*
Within (W)	253	26052.1	3445.1	3183.2	2727.62	252	10.82	
Total (T)	254	27632.1	4127.6	3478.0	2861.42	253		
Physical Causal Test Part I (Fifth-grade Ss)								
Methods (A)	1	1252.0	798.0	508.6	206.60	1	206.60	11.68*
Within (W)	156	21221.9	4654.5	3762.3	2741.51	155	17.69	
Total (T)	157	22473.9	5452.4	4270.9	2948.11	156		
Physical Causal Test Part II (Fifth-grade Ss)								
Methods (A)	1	1252.0	213.5	36.4	2.05	1	2.05	.18
Within (W)	156	21221.9	2734.6	2132.9	1780.56	155	11.49	
Total (T)	157	22473.9	2948.1	2169.3	1782.61	156		

* Significant at the .001 level

TABLE IX

COMPARISON OF THE MEAN SCORES OF THE EXPERIMENTAL AND CONTROL FIFTH-GRADE SUBJECTS ON TWO SUBSCORES OF THE PHYSICAL CAUSAL TEST (1958-59 DATA)

	Experimental Ss			Control Ss			
	N	Mean	SD	N	Mean	SD	t
Pre-Test Fall 1958 Physical Causal Test Form A Part I	52	8.50	2.55	65	9.15	3.02	1.26
Growth Fall-Spring Physical Causal Test Form B I - A I	52	2.08	2.62	65	.58	2.59	3.05*
Pre-Test Fall 1958 Physical Causal Test Form A Part II	52	3.79	1.82	65	4.12	1.95	.95
Growth Fall-Spring Physical Causal Test Form B II - A II	52	.81	1.98	65	.83	2.50	.05

*Significant at the .005 level

kind of problems which appear in PCT Part I but not those in PCT Part II. These explanations are speculative, and there may be others.

Nevertheless, it is obvious that the nature of the items in PCT Part I is basically different from those in Part II. This is also indicated by the moderately high correlation between these two parts of the Physical Causal Test. The correlation between the PCT Part I and PCT Part II subscores is .42 (N = 158) for fifth-grade Ss and .44 (N = 245) for sixth-grade Ss, indicating that the two parts of the PCT measure two different aspects of the S's understanding of his physical environment, since the amount of variance that can be accounted for by the correlation of the two measures is 20 percent.

The question arises when the 1957-58 study was repeated for fifth-grade Ss in 1958-59 whether the findings were duplicated. Inspection of Table IX supports the findings from the preceeding year, even though growth measures are utilized. On the pre-test scores of both PCT Part I and PCT Part II there are no significant differences between experimental and control Ss. On the growth measure, obtained by subtracting the pre-score from the post-score, there is a significant difference on PCT Part I but not on PCT Part II. This is an exact duplication of the 1957-58 data, even though the revised PCT test contained only half as many Part I and Part II items. The assumption that the increase in score is due to the experimental learning program is substantiated by the fact that in the 1958-59 data there existed no known differences on the criterion variables at the beginning of the experiment.

These findings show that on these two subtests of the Physical Causal Test the learning effect of the experimental program affects fifth- and sixth-grade Ss differently. Further investigation is needed to satisfactorily explain these findings.

Summary

Experimental (N = 208) and control (N = 205) Ss from fifth- and sixth-grade classes were compared in respect to their mean scores on measurements of social and physical causality in order to determine the effect and the transfer effect of a mental health learning program which is designed to develop an understanding and appreciation of the motives of human behavior and an awareness of the dynamics of factors operating in social situations. Data were analyzed using a co-variance design in order to eliminate the influence of IQ on the findings.

The findings of the study allow for the following generalizations:

1. The experimental learning program substantially contributes to an understanding of the dynamic, complex, and variable nature of human behavior in fifth- and sixth-grade Ss as measured by the ECT.
2. The learning program reduces a fifth-grade S's willingness to be punitive as measured by the PST. Whether it does so for sixth-grade Ss cannot be assessed clearly due to certain limitations of the test instrument.
3. The learning program has a transfer effect for fifth- and sixth-grade Ss on an understanding of the causes of common events and phenomena in the physical world and an awareness of the probabilistic nature of knowledge as measured by the PCT.
4. The transfer effect is different for fifth- and sixth-grade Ss. Fifth-grade experimental Ss have a better understanding of the factors that produce a common event, as measured by the PCT Part I, than their control Ss. Sixth-grade experimental Ss have a better understanding of the probabilistic nature of knowledge, as measured by the PCT Part II, than their respective control Ss. Further investigation is needed to explain this phenomenon.
5. Measures of physical causality have a higher correlation with IQ than measures of social causality.
6. There is a tendency for experimental Ss to obtain lower correlations between the measures of causality and IQ than is the case for control Ss. It appears that the learning program increases a S's understanding of social and physical causality regardless of the S's level of intellectual functioning.
7. There is a developmental increase in understanding of causality as measured by the ECT, PST and PCT from fifth- to sixth-grade for control Ss. This increase appears to be greater in the physical than in the social science area.
8. Using fifth-grade Ss and modified test instruments in the subsequent year, most findings were duplicated. Since it could be demonstrated that no known differences existed on the criterion variables at the beginning of the experiment, the follow-up study provides substantial evidence for the assumption that these effects are due to the experimental learning program.
9. While one cannot draw conclusions beyond this particular experimental learning program and beyond the specific test instruments utilized in this study, the findings are supported in general by other studies conducted as part of the Preventive Psychiatry Program (10, 13, 14, 16, 20).

REFERENCES

1. Clark, C. A. *A Study of the Adolescent's Conception of the Nature of Scientific Knowledge*, Ph. D. dissertation, State University of Iowa, 1954.
2. Deutsche, Jean M. "The Development of Children's Concepts of Causal Relations," in *Child Behavior and Development* (New York: McGraw-Hill Book Co., 1943), pp. 129-45.
3. Fitzgerald, D. C. and Ojemann, R. H. "A Study of the Growth of the Dynamic Concept of Knowledge," *Child Development*, XV (1944), pp. 53-62.
4. Huang, I. "Children's Conception of Physical Causality: A Critical Summary," *Journal of Genetic Psychology*, LXIII (1943), pp. 71-121.
5. Judd, C. H. "The Relation of Special Training to General Intelligence," *Educational Review*, XXXVI (1908), pp. 28-42.
6. Lacey, J. I. and Dallenbach, K. M. "Acquisition by Children of the Cause-Effect Relationship," *American Journal of Psychology*, LII (1939), pp. 103-10.
7. Levitt, E. E. "Punitiveness and 'Causality' in Grade School Children," *Journal of Educational Psychology*, XLVI (1955), pp. 494-98.
8. Levitt, E. E. and Ojemann, R. H. "The Aims of Preventive Psychiatry and 'Causality' as a Personality Pattern," *Journal of Psychology*, XXXVI (1953), pp. 393-400.
9. Lyle, W. H. and Levitt, E. E. "Punitiveness, Authoritarianism, and Parental Discipline of Grade School Children," *Journal of Abnormal and Social Psychology*, LI (1955), pp. 42-46.
10. McNiel, Bessie F. "Development at the Youth Level of a Conception of the Causes of Behavior and the Effectiveness of a Learning Program in this Area," *Journal of Experimental Education*, XIII (1944), pp. 81-85.
11. Muuss, R. E. "Mental Health Effects of a One- and Two-Year Causal Learning Program on Sixth Grade Subjects when IQ, Age, and Sex are Controlled," *American Psychologist*, XV (1960), p. 435.
12. Muuss, R. E. "Mental Health Implications of a Preventive Psychiatry Program in the Light of Research Findings," *Marriage and Family Living*, XXII (1960), pp. 150-56.
13. Muuss, R. E. "The Relationship Between 'Causal' Orientation, Anxiety, and Insecurity in Elementary School Children," *Journal of Educational Psychology*, LI (1960), pp. 122-29.
14. Ojemann, R. H. "The Effect on the Child's Development of Changes in Cultural Influences," *Journal of Educational Research*, XL (1946), pp. 258-70.
15. Ojemann, R. H. "A Functional Analysis of Child Development Material in Current Newspapers and Magazines," *Child Development*, XIX (1948), pp. 76-92.
16. Ojemann, R. H. and others. "The Effects of a 'Causal' Teacher-Training Program and Certain Curricular Changes on Grade School Children," *Journal of Experimental Education*, XXIV (1955), pp. 95-114.
17. Ojemann, R. H. and Snider, B. "An Approach to the Teaching of Human Relations," *Journal of Teacher Education*, X (1959), pp. 235-40.
18. Piaget, J. *The Child's Conception of Physical Causality*, translated by M. Gabain (New York: Harcourt Brace and Co., 1930).
19. Stiles, Frances S. *A Study of Materials and Programs for Developing an Understanding of Behavior at the Elementary School Level*, Ph. D. dissertation, State University of Iowa, 1947.
20. Stiles, Frances S. "Developing an Understanding of Human Behavior at the Elementary Level," *Journal of Educational Research*, XLIII (1950), pp. 516-24.
21. Vinacke, W. E. "Concept Formation in Children of School Age," *Education*, LXXIV (1954), pp. 527-534.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 3, March 1961)

A COMPARISON OF SCORES OBTAINED BY ADMINISTERING A TEST NORMALLY AND VISUALLY *

H. A. CURTIS and R. P. KROPP
Florida State University

THIS IS A REPORT of the first of a series of studies designed, in part, to ascertain the feasibility of administering via television group standardized tests which are commonly used in the public schools. Since an increasing amount of instruction is executed through the medium of television, it is relevant to inquire whether testing might also be done by television in such a way as to obtain, on the one hand, the same results for students as would be obtained had they been tested under normal conditions, and, on the other hand, to obtain more valid results than are afforded by the usual method of administration. Also, it is relevant to explore the possibilities afforded by television with respect to gaining test data that simply cannot be obtained by the traditional paper-and-pencil tests due to their inherent limitations. This report deals only with a description of attempts to gain by television administration test results which are comparable to those obtained under normal administrative conditions.

An examination of television as a device by which to communicate test materials revealed a number of problems. The most notable of these problems is the precise method by which the test might be presented within the limiting conditions imposed by the nature of television. When a test is administered under normal conditions, it is usual to inform the examinee of the total time allowed and then to present him with all the test items. He is free to determine the amount of time he will spend on each item, to select items he will attempt, and to re-attempt the items which he had previously omitted. But the size of the television screen precludes presenting all items simultaneously to the student, thus some of the freedoms permitted in the normal administration of tests will be denied him when the test is presented by television. Consequently, a substantially different test administration technique must be used for television testing.

The technique which is reported here is called "pacing." It involves the serial presentation of one test item or a small group of items at a time until all items in the test have been exhausted. It does not allow the examinee to refer back to items which have previously been exposed. A single presentation of an item or group of items is referred to as an exposure. The duration of the exposure is called the intra-exposure interval. The time elapsing between the end of one exposure and the beginning of the next exposure is called the inter-exposure interval. The serial exposure of small portions of tests, with the exposures possibly being separated by noticeable intervals of time, makes possible presentation of test materials by television. However, it is radically different from normal procedure and it poses new variables in the administrative technique which must be studied.

In addition to the variables already mentioned, there are at least the following: item-order, item-type, audio and/or visual presentation, kind of test (intelligence, achievement, adjustment, etc.), guessing instructions, etc.

The possibility of using the pacing technique hinges on the possibility of setting values on these variables that will have the effect of producing test results comparable to those that would have been obtained under normal administrative conditions. The following is a description of a first-attempt in this direction.

Problem

The general problem reported on here was a study of the changes in test scores of a group of subjects who were administered the same test under three sets of administrative conditions. These three sets of conditions are as follows. One administration of the test was done in conformance with normal procedures, i.e., pupils were

---

*Footnotes will be found at the end of the article.

given test booklets, answer sheets, etc. This administration is subsequently referred to as the "control" administration. The other two sets of administrative conditions were experimental in nature and involved projecting test items on a screen as the method of presenting the test to the subjects. One of these experimental administrations consisted of presenting test items one at a time to the examinees and it is referred to as the "single" presentation. The other experimental administration involved presenting the items three at a time and is referred to as the "triad" administration. More explicit descriptions of the experimental conditions appear in the subsequent section.

The following comparisons were made of the scores obtained under these three sets of conditions: means, frequency of response, guessing, and the factor patterns of each set of scores.

Although the over-riding purpose is to determine a feasible method of administering tests by television, the experimental administrations of the test were not done by television due to the expense and awkwardness which it would have entailed. Items were reproduced on slides and were projected on a 3'x 5' screen. Slide format conformed to the format of the TV screen. The illumination level maintained in the testing room was comparable to that maintained for TV viewing.

Procedures

Sample. The sample consisted of a ninth-grade class of white students. Complete sets of test data were collected from twenty-nine of the class members. Four students from whom incomplete test data were gathered were discarded prior to the analysis of the data.

Tests. The following tests were administered to the subjects under normal conditions:

School Ability Test, Form 3A, of the Cooperative School and College Ability Test[2]

Iowa Test of Educational Development, Form X3S[3], Tests 3-7

Test 3. Correctness and Appropriateness of Expression

Test 4. Ability to do Quantitative Thinking

Test 5. Ability to Interpret Reading Materials in the Social Sciences

Test 6. Ability to Interpret Reading Materials in the Natural Sciences

Test 7. Ability to Interpret Literary Materials

Thurstone Temperament Schedule[3]

Gordon Personal Profile[4]

Gordon Personal Inventory[4]

SRA Primary Mental Abilities, ages 11-17[3]

Iowa Silent Reading Test, New Edition, Advanced[4]

Clerical Speed and Accuracy of the Differential Aptitude Test[5]

The School Ability Test was administered normally and experimentally. All other tests mentioned above were administered to provide data on the basis of which to evaluate the differences between the results obtained from the normal and experimental administrations of the School Ability Test.

The School Ability Test was used as the experimental test[6]. The test consists of four parts. Parts I and III assess verbal abilities, and Parts II and IV assess quantitative abilities. On the basis of item analysis data supplied by the Educational Testing Service, the test was broken into two half-tests, thus permitting the use of a half-test for each set of experimental conditions and thereby eliminating item overlap between the two experimental administrations. Data dealing with the comparability of these half-tests appear in Table I. The column entitled "N" indicates the number of items in each part of the half-test.

Experimental Conditions. There were similarities and differences between the experimental conditions under which the half-tests were administered on two occasions. The similar conditions were as follows: no inter-exposure interval, all items were five-choice multiple choice, the total testing time devoted to item solution was the same, directions were the same, and the mode of item presentation was visual.

The chief differences between the experimental administrations were the number of items presented per exposure, the intra-exposure interval, and the order in which the items were presented.

The single item administration presented one item per exposure. The exposure duration was determined by dividing the number of items included in each part of the test into the amount of time the test authors allotted for that part of the test. Since the number of items and time allowed differed from part to part of the test, the exposure interval was not constant throughout. The following durations were used: Part I, 30 seconds; Part II, 48 seconds; Part III, 20 seconds; and Part IV, 60 seconds. The items within each part of the half-test were presented in the same order in which they appeared in the parent test, thus the item order was that of ascending difficulty.

The triad administration presented three items per exposure. The exposure interval for each part was three times that allowed for the single item administration. The item order within each part was determined in this way. Triads were constructed to have the same mean difficulties and ranges, insofar as possible. Within a triad, items were arranged from least to most difficult. Since it was impossible to make the range of difficulty for each triad exactly the same, the triads were arranged from greatest to least range. In short, the order of items within a triad presented a rhythm of difficulty--simple, average, difficult --which was repeated throughout the entire test.

TABLE I

COMPARISON OF HALF-TESTS DEVISED FROM THE SCHOOL ABILITY TEST

	Form A				Form B		
Part	$\bar{p}$	Sigma	p range	N	$\bar{p}$	Sigma	p range
I	.48	.24	.14-.91	15	.48	.20	.07-.75
II	.47	.24	.09-.91	13	.47	.23	.09-.84
III	.45	.21	.11-.82	15	.45	.23	.14-.95
IV	.44	.23	.06-.80	13	.44	.24	.08-.80

TABLE II

MEANS AND STANDARD DEVIATIONS ON HALF-TESTS FROM SAT, 3A

	Mean	S.D.
CVS	17.068	5.98
CVT	16.757	6.00
CQS	15.619	3.62
CQT	14.930	3.86
XVS	17.930	6.05
XVT	18.102	6.92
XQS	15.378	3.64
XQT	15.550	3.74

N = 29

The reader will note that thirteen items are contained in each of Parts II and IV of the test. Actually fifteen items were used in each so that there would be five triads. Two were borrowed from the comparable part of the other form. However, these borrowed items were not scored.

Testing Schedule. The tests were administered in two sittings, the first preceding the second by a period of three months. The first sitting included the normal administration of the School Ability Test and the five tests of the Iowa Tests of Educational Development. These tests were administered in half-day sessions on three consecutive days.

The remaining tests were administered over a three-day period as follows: First day, morning --Gordon Personal Profile and Gordon Personal Inventory; afternoon--single item experimental half-test. Second day, morning--SRA Primary Mental Abilities; afternoon--triad experimental half-test. Third day, morning--Thurstone Temperament Schedule, DAT Clerical Speed and Accuracy, and Test I of the Iowa Silent Reading Test.

Experimental Directions. Special directions and sample items were used when administering the experimental test. The directions were distributed, one copy to each subject, and were read aloud by the examiner while the subjects read them silently. Prior to the presentation of each of the four parts of each experimental test, special directions were given and a sample item was exposed on the screen. Subjects were at no time informed of the amount of time they could spend on each exposure. They were informed that no questions could be asked during the administration of any part of the test. They were informed in the instructions that they would not be able to refer back to items which they might not have answered when they had been presented.

No unusual circumstances arose during the administration of the experimental tests.

Results

Comparison of Means. The School Ability Test that was administered under normal conditions was re-scored several times to gain scores over items which appeared in the half-tests. In accordance with customary usage of the School Ability Test, Parts I and III were combined to provide a verbal score and Parts II and IV were combined to provide a quantitative score.

The means and standard deviations are presented in Table II on the previous page. In this table, and hereafter, the following code is employed to conserve space: C, control; X, experimental; V, verbal; Q, quantitative; S, single; and T, triad.

Because all sets of data were gathered from the same subjects, it was necessary when computing t tests of mean differences to make corrections for correlated performances. The relevant intercorrelations appear in Table III and the t tests appear in Table IV.

The following generalization might be tentatively drawn from the tabular data. The XT administrative conditions lead to higher scores than do the XS conditions. On the verbal section, either experimental condition produces higher mean scores than does the control administration. But apparently no generalization can be drawn for the quantitative items: CQS and CQT means bracket the counterpart X means despite the fact that the CQS and CQT means should be equal on account of the method used to build the half-tests.

The tendencies described above must be regarded skeptically because the t tests revealed no significant differences between the means except in one case. Furthermore, one might attribute the apparent, although not significant, superiority of XT to XS to the fact that XT data were gathered subsequent to the XS data thereby reflecting practice effect.

In summary, the data to this point do not generally indicate, at the .05 level, differences in mean scores which are functions of administrative conditions. However, the control data were collected three months prior to the collection of experimental data. Nevertheless, the intercorrelations which appear in Table III indicate that subjects would fall in roughly the same order on the experimental test as they were ordered by the control test. Thus, the data to this point would, in a limited way, lend support to the administration of tests by television if a pacing technique is used.

Comparisons by Item Types. The preceding analysis dealt with means on the Q and V sections of the test inasmuch as Q and V scores are generally reported when the School Ability Test is used. However, the Q and V scores, respectively, are established by summing success over two parts of the test each of which has relatively distinct kinds of items. Sample items illustrate this difference.

Verbal

Part I We had worked hard all day so that by evening we were quite ( ).
A. small B. tired C. old
D. untrained E. intelligent

Part III Chilly
A. tired B. nice C. dry
D. cold E. sunny

Thus, the items in Part I contain a sentence as a stem, whereas in Part III the item stem is merely a stimulus word.

Quantitative

Part II 5413 A. 586 B. 596 C. 696
-4827 D. 1586 E. None of these

TABLE III

CORRELATIONS BETWEEN SCORES ON CONTROL AND EXPERIMENTAL HALF-TESTS OF SAT, 3A

	CVS	CVT	XVS	XVT	CQS	CQT	XQS	XQT
CVS		.879	.935	.864				
CVT			.869	.865				
XVS				.854				
XVT								
CQS						.815	.789	.687
CQT							.650	.703
XQS								.753
XQT								

N = 29

TABLE IV

"t" TEST OF MEAN DIFFERENCES CORRECTED FOR CORRELATED PERFORMANCES

	CVS	CVT	XVS	XVT	CQS	CQT	XQS	XQT
CVS		.55	-2.03	-1.59				
CVT			-1.99	-2.07*				
XVS				- .25				
XVT								
CQS						1.60	.53	.12
CQT							-.76	-.73
XQS								-.25
XQT								

*p .05

N = 29

Part IV Four $10-bills are equal to how many $5 -bills?

A. 20 B. 10 C. 8
D. 40 E. 2

Thus, the items in Part II consist of stems containing a set of numbers and an operations sign, whereas the items in Part IV consist of stems which are statements of word problems involving arithmetic.

Analyses were done to determine whether differences existed between success on particular item types in association with different conditions under which the items were administered. The relevant data appear in Tables V (means and standard deviations), VI (intercorrelations) and VII (t's) on the next pages.

The tabular data reveal for the items in the two verbal parts of the test that regardless of whether the item contained a verbal statement or not the experimental administrative procedures led to slightly higher mean scores than did the normal administrative conditions. Whether this is due to the effect of the experimental conditions or the fact that the experimental data were collected three months after the control data is very much in doubt. On the quantitative items a differential in response accuracy is associated with method of test administration. On the quantitative items having statements for stems, the experimental administrations led to higher mean scores than did the control administration. However, on the quantitative items which had for a stem a series of numbers and an operation sign, the control means were significantly different from and greater than the experimental means. This finding is worthy of comment.

It is felt that the reduced mean scores on the quantitative no-stem items is attributable to any or all of the following, since the problems were so involved the examinee probably copied the data from the screen to his scratch paper in order to solve it, thus 1) reducing his working time due to the time involved in copying, 2) committing clerical errors when copying, 3) after solving the problem, needing to locate a similar response on the screen during which errors in remembering might have occurred. Only further research will provide definite knowledge of the reasons for the reduced scores. It should be noted at this time, though, that non-response to these items was significantly less than the amount of non-response on the same items when administered normally. Whether the poorer performance is due simply to guessing or to factors noted above is problematical.

Another factor worthy of specific mention is that the intercorrelations of the quantitative items with regard to method of administration are of a lesser magnitude than comparable intercorrelations for the verbal items. Generally the lowest intercorrelations are based on the no-statement quantitative items.

Analysis of Omits. It was hypothesized that subjects would be more likely to give a greater number of responses to the items when experimentally presented than when normally presented simply because all items would be exposed to the examinees on the experimental administration whereas on the normal administration their test taking habits might have caused them to omit some items. Also it was hypothesized that there would be no difference in response frequency between the triad and single experimental presentations. Both of these hypotheses held at the .05 level of confidence. The Wilcoxon Matched-Pairs Signed-Ranks Test was used in computing the significance of differences. The pertinent data are presented in Table VIII.

Thus it appears that the experimental test administration procedures, i.e., pacing examinees over all items, leads to a greater number of item responses than does the normal administrative procedure during which the examinees are responsible for their own pacing.

Although differences in non-response are associated with mode of test administration, the data were analyzed to determine if non-response was related to certain personality variables on one kind of administration but not on the other. A response accuracy score was calculated for each examinee on the basis of his responses on the XQS and CQS administrations, separately. The accuracy score for a subject was equal to the ratio of his "rights" score to the total number of items he responded to. The score was computed in this way to discount differences in ability which would effect the subject's probability of non-response. The personality variables which were evaluated with regard to response accuracy were "Cautiousness" from the Gordon Personal Inventory and "Reflective" and "Impulsive" from the Thurstone Temperament Schedule. A 2 x 2 χ^2 analysis produced the results shown below.

CQS response accuracy		
vs Cautious	$\chi^2 = 00$	$p > .05$
vs Reflective	$\chi^2 = 2.28$	$p > .05$
vs Impulsive	$\chi^2 = .56$	$p > .05$
XQS response accuracy		
vs Cautious	$\chi^2 = .65$	$p > .05$
vs Reflective	$\chi^2 = .00$	$p > .05$
vs Impulsive	$\chi^2 = .00$	$p > .05$

The data indicate that tendency to omit items is not related to these three personality variables on either the normal or experimental test administrations.

Intercorrelations. The intercorrelations between control and experimental administrations of the experimental tests appear in Table IX. The

TABLE V

MEANS AND STANDARD DEVIATIONS OF PART SCORES ON HALF-TESTS FROM SCHOOL ABILITY TEST

Condition	Verbal				Condition	Quantitative			
	S*(I)		NS**(III)			S(IV)		NS(II)	
	X	sigma	X	sigma		X	sigma	X	sigma
CVS	8.52	3.89	8.55	2.94	CQS	6.48	2.01	9.14	2.11
XVS	9.24	2.98	8.69	3.28	XQS	7.24	1.79	8.10	2.26
CVT	8.45	2.96	8.31	3.46	CQT	6.45	2.01	8.48	1.45
XVT	8.79	3.97	9.31	3.51	XQT	8.24	2.18	7.31	2.37

N = 29

* stem

**no stem

TABLE VI

CORRELATIONS BETWEEN PART SCORES ON HALF-TESTS FROM SCHOOL ABILITY TEST

Variable Pair	Verbal S(I) r	Verbal NS(III) r	Variable Pair	Quantitative S(IV) r	Quantitative NS(II) r
CVS-XVS	.77	.86	CQS-XQS	.58	.71
CVT-XVT	.76	.79	CQT-XQT	.59	.48
XVS-XVT	.71	.76	XQS-XQT	.54	.41

N = 29

TABLE VII

"t" TESTS OF DIFFERENCES BETWEEN CORRELATED MEANS OF PART SCORES FROM HALF-TESTS OF SCHOOL ABILITY TEST

Variable Pair	Verbal S(I) t	Verbal NS(III) t	Variable Pair	Quantitative S(IV) t	Quantitative NS(II) t
CVS-XVS	-1.55	- .43	CQS-XQS	-2.29*	3.28*
CVT-XVT	- .71	-2.36*	CQT-XQT	-5.03*	3.00*
XVS-XVT	.846	-1.40	XQS-XQT	-2.73*	1.67

*$p < .05$

N = 29

TABLE X

ROTATED FACTOR MATRIX*

Tests	I	II	III	IV	V	VI	VII	VIII
01 CVS	9470	-0593	-0261	0279	-0622	0450	1028	-0497
02 CQS	7705	0607	-0826	-0001	-0741	-5216	1367	-1388
03 CVT	9382	0903	-0619	-0032	-1244	0695	0778	0302
04 CQT	6866	-0693	-1172	0164	-0942	-5056	-0170	2789
05 Control Total Score	9616	-0272	-0680	-0206	-0950	-1538	0797	0447
06 XVS	9172	0117	0096	1013	-0833	1002	2484	0740
07 XQS	6916	-2285	1308	-1153	0030	-5205	0192	-1300
08 XVT	8979	0640	1574	-0131	-0206	1715	0254	0266
09 XQT	6983	-0933	0288	-2227	0233	-3572	-0874	1214
10 Gordon Ascendancy	1945	-8023	-1591	0575	-0143	-0468	0624	-0155
11 Gordon Responsibility	2244	-0709	7513	-1007	-1000	0539	2756	-0272
12 Gordon Emotionally Stable	-0769	2919	5357	-0118	0053	1544	3273	0488
13 Gordon Sociability	-0349	-8326	0611	0615	0541	-2651	0428	-1051
14 Gordon Cautiousness	0177	-1044	8164	-0262	1668	0581	-1055	0201
15 Gordon Original Thinking	3424	-3108	6156	0625	-1326	-0842	1408	0299
16 Gordon Personal Relation	0506	1643	7077	0675	0727	-1656	-3099	-0778
17 Gordon Vigor	3675	-1835	2388	-1200	-1771	0848	1920	-3291

18 *Thurstone Active*	2162	-5710	-2405	1607	-3380	-0101	0045	-1353
19 Thurstone Vigorous	0141	-2368	0327	6066	0033	2479	-0151	0150
20 Thurstone Impulsive	-0334	-5266	1619	-1468	-0926	0039	4332	-1165
21 Thurstone Dominant	0551	-8274	-0373	-2266	0864	1955	-1445	1761
22 Thurstone Stable	0084	-3282	2241	-6277	-0572	1103	-0068	-0376
23 Thurstone Sociable	-2000	-8546	2144	0256	-0628	0359	-0068	0164
24 Thurstone Reflective	1779	-0469	0887	3048	1842	1023	1308	4205
25 PMA Verbal Meaning	8326	1297	0700	2128	2912	1519	0246	-0742
26 PMA Space	3468	-0397	-4441	2464	1892	1527	2155	1830
27 PMA Reasoning	4617	0644	-1080	-3836	4377	0240	-0485	1625
28 PMA Number	6178	-2343	-3273	-0321	1393	1792	1419	-3654
29 PMA Word Fluency	6544	-0806	2111	0169	0016	-1743	1670	-3978
30 PMA Scholastic Aptitude	8401	0214	0320	1102	4630	1406	0120	-0016
31 DAT Clerical Sp. and Acc.	3204	-0672	0410	1342	6190	0276	-0068	-0371
32 Iowa Silent Reading Rate	5785	-0459	0503	3930	1196	0843	-2257	1481
33 Iowa Reading Comprehension	8074	1520	0140	0178	1090	-2649	0221	1216
34 ITED Correct Expression	8864	1020	1034	-2423	0852	1187	0409	-0125
35 ITED Quantitative	7562	-1799	-0538	0121	-0746	-3008	-1130	-1024
36 ITED Social Studies	8527	0642	0200	-0307	-0138	0829	-4265	0980
37 ITED Natural Science	8050	0542	1314	-0053	-0844	0956	-3647	-1054
38 ITED Literary Material	8681	0084	-0433	-0720	-0921	1863	-3272	0304

---

* All decimal points omitted.

principles of television testing.

FOOTNOTES

1. This investigation was financed by grants from the Graduate Research Council of the Florida State University, and from the U.S. Department of Health, Education, and Welfare, Office of Education, Title VII, National Defense Education Act of 1958, Contract Number 7-08-075.00.
2. Educational Testing Service, Princeton New Jersey.
3. Science Research Associates, Chicago, Illinois.
4. World Book Company, Tarrytown, New York.
5. Psychological Corporation, New York.
6. The authors are indebted to the Educational Testing Service for permission to use this test experimentally.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 3, March 1961)

NONNORMALITY AND PRODUCT MOMENT CORRELATION*

RAYMOND C. NORRIS
George Peabody College for Teachers
HOWARD F. HJELM
United States Office of Education

ALTHOUGH THE Pearson product moment correlation coefficient has been used as an index of relationship between two variables for more than 50 years, there continues to be considerable controversy concerning the circumstances under which its use is appropriate. As recently as 1957 and 1958 The American Psychologist carried comments by Nefzger and Drasgow (8), Furfey (5), LaForge (6), and Milholland (7) on the question of whether or not it is necessary to assume normality in the distributions on which the correlations are based. Binder, in a more recent article in The American Psychologist (2), attempted to show that much of the controversy has been misdirected because of a lack of understanding of the place of assumptions in the interpretation of the correlation coefficient. By developing three different mathematical models, Binder shows that different inferential interpretations of the correlation coefficient grow out of the different mathematical models and concludes, "The preference for the bivariate normal model to a more general model stems from its great deductive power and its usefulness in many empirical situations. . . ."

The extent to which failure to satisfy the assumption of bivariate normality distorts the interpretation of the product moment coefficient was the subject of numerous studies in the period 1929 through 1932. Egon Pearson (9, 10, 11) conducted a series of studies in which the score distributions deviated from bivariate normal in various ways and concluded that the sampling distribution of the product moment coefficient was not seriously effected by failure to meet this theoretical assumption. His findings were generally confirmed by other empirical investigations by Dunlap (4), Chesire, Oldis, and Pearson (3) and Rider (12). Baker (1), dealing with a bivariate distribution with markedly skewed marginal frequencies, concluded that for samples of size forty the obtained sampling distribution was also skewed and that usual tests of significance would be misleading.

Marginal distributions in these studies deviated from the normal in both skewness and kurtosis. Population correlations ranged from zero to .83, sample sizes from 5 to 52, and sampling distributions from 50 coefficients to 1770. In most of the work sampling distributions were based on 500 or fewer correlations, and Baker's study (the only one which raised serious question about the distortion of the obtained sampling distribution) was based on only 50 samples of size 40.

The current study was undertaken in an attempt to pin down any empirical effects of nonnormality by taking larger numbers of samples than had been used previously and determining the obtained sampling distribution in each case. Sample sizes used were in the range of ones frequently encountered in educational and psychological research. Several types of nonnormality frequently encountered in such research were used, and the population correlations of zero and .83 represented the range of values usually encountered.

Procedure

Ten populations of size ten thousand were established. A population having approximately no correlation and a population having substantial correlation were established for each of the following bivariate forms: normal, rectangular, leptokurtic, slightly skewed, and markedly skewed. The ten populations were punched onto one set of ten thousand IBM cards. The ten thousand cards were placed in random order by ordering four columns of random digits. The samples were obtained by repeatedly counting off the desired number of cards. The deck was sampled without replacement and to exhaustion. The deck was reordered by using another set of four columns of random digits and samples counted off. This was repeated twenty-four times in order to obtain the desired number of samples. This necessitated having five decks of ten thousand cards each with the sixty columns of data being identical for each deck leaving a total of one hundred columns for random digits. Sample sizes 15, 30, and 90 were obtained concurrently with the first two samples of size 15 comprising the first sample of size 30, and the first three samples of size 30 comprising the first sample of size 90. Correlation coefficients were calculated for the samples and sampling distributions for sample sizes 15, 30, and 90 were established for each of the ten populations. As indicated in Table I, the sampling distributions consisted of 15,984 coefficients for sample size 15; 7,992 coefficients for sample size 30, and 2,664 coefficients for sample size 90. The sampling and calculating were done by means of the IBM 650 Data Processing System.

---

*The research reported herein was performed pursuant to a contract with the Office of Education, United States Department of Health, Education and Welfare.

The obtained sampling distributions were compared with the appropriate theoretical sampling distributions assuming bivariate normality and with the obtained sampling distributions from the bivariate normal populations. Fisher's transformations,

$$z = \tfrac{1}{2}\log_e \frac{1 + r}{1 - r} \quad ,$$

were used in establishing the theoretical sampling distributions. The obtained and theoretical sampling distributions were tested for overall goodness of fit by means of Kolmogorov-Smirnov tests. Critical values were established for $r_{.005}$, $r_{.025}$, $r_{.975}$, and $r_{.995}$ and frequency tabulations made of the number of significant coefficients in each of the obtained sampling distributions.

Results

The obtained sampling distributions from the normal (control) population having no correlation, approximated its theoretical distribution very closely. The Kolmogorov-Smirnov goodness-of-fit test showed no departure from the theoretical distribution for samples of sizes 30 and 90, and the discrepancy between obtained and theoretical distributions for sample size 15 was significant only at the .20 level of significance. The closeness of the fit even in this case was remarkable. Inspection of Table II and Graph 1 will also show that for the normal case the relative incidence of significant correlations conformed very closely to the theoretical proportions.

The close correspondence of the obtained and theoretical sampling distributions where the population correlation was approximately zero and the score distribution approximately normal was interpreted as providing assurance of the adequacy of the sampling procedures used in this study. In particular, the effects of using discrete, finite populations in place of the theoretical continuous and infinite populations did not appear to materially influence the results. Nor did sampling procedures, as such, appear to cause deviation of the empirically determined distributions from the theoretical expectations.

The empirical sampling distributions based on a rectangular population conformed very closely to their theoretical distributions. However, the Kolmogorov-Smirnov test indicated a significant departure between the two for sample size 15. Although the probability of such a discrepancy occurring by chance alone was less than .05, the magnitude of the discrepancy was small. Inspection of Table II and Graph 2 will also show that the incidence of significant correlations did not deviate markedly from the theoretical proportions at either the .05 or .01 levels of significance. All considered, there was little evidence that sampling from a rectangular distribution produced sampling distributions markedly different from those obtained when the assumption of bivariate normality was met.

The obtained sampling distributions based on a leptokurtic population deviated from their theoretical distributions in the same fashion as did those based on the normal population, but the magnitued of the discrepancies was much greater. The differences were significant at the .05 level for samples of size 15 and 90 (see Graph 13). For samples of size 90 significant correlations were obtained only .69 as often as expected at the .05 level of significance and .60 as often as expected at the .01 level. Although there was some evidence that the sampling distributions of correlation coefficients based on a leptokurtic population tended to be leptokurtic also, the departure from the theoretical distribution was not very great.

The obtained sampling distributions based on slightly skewed or markedly skewed populations appeared to bear very similar relationships to their theoretical distributions. Kolmogorov-Smirnov tests on the six comparisons indicated no significant departure from the theoretical distributions at the .20 level of significance except for samples of size 15 based on the markedly skewed population. The departure in that case was one which could occur between ten and fifteen percent of the time by chance alone. Reference to Table II and Graph 5 shows that even here the relative incidence of significant coefficients did not deviate markedly from theoretical expectations. Only in the fact that there was a tendency to have too few significant negative correlations and too many significant positive correlations was there any suggestion that sampling distributions based on skewed populations would be skewed also.

In general, it was concluded that for populations in which there was no correlation the sampling distributions based on nonnormal populations did not differ markedly from those where the theoretical assumptions of normality could be met. For samples as small as 15 or as large as 90, the critical regions contained approximately the correct proportion of extreme coefficients at either the .05 or .01 levels of significance. Use of the theoretical distribution in place of some more exact sampling distribution would not seriously influence the probability of making errors of the first kind.

Sampling distributions based on samples of

TABLE I

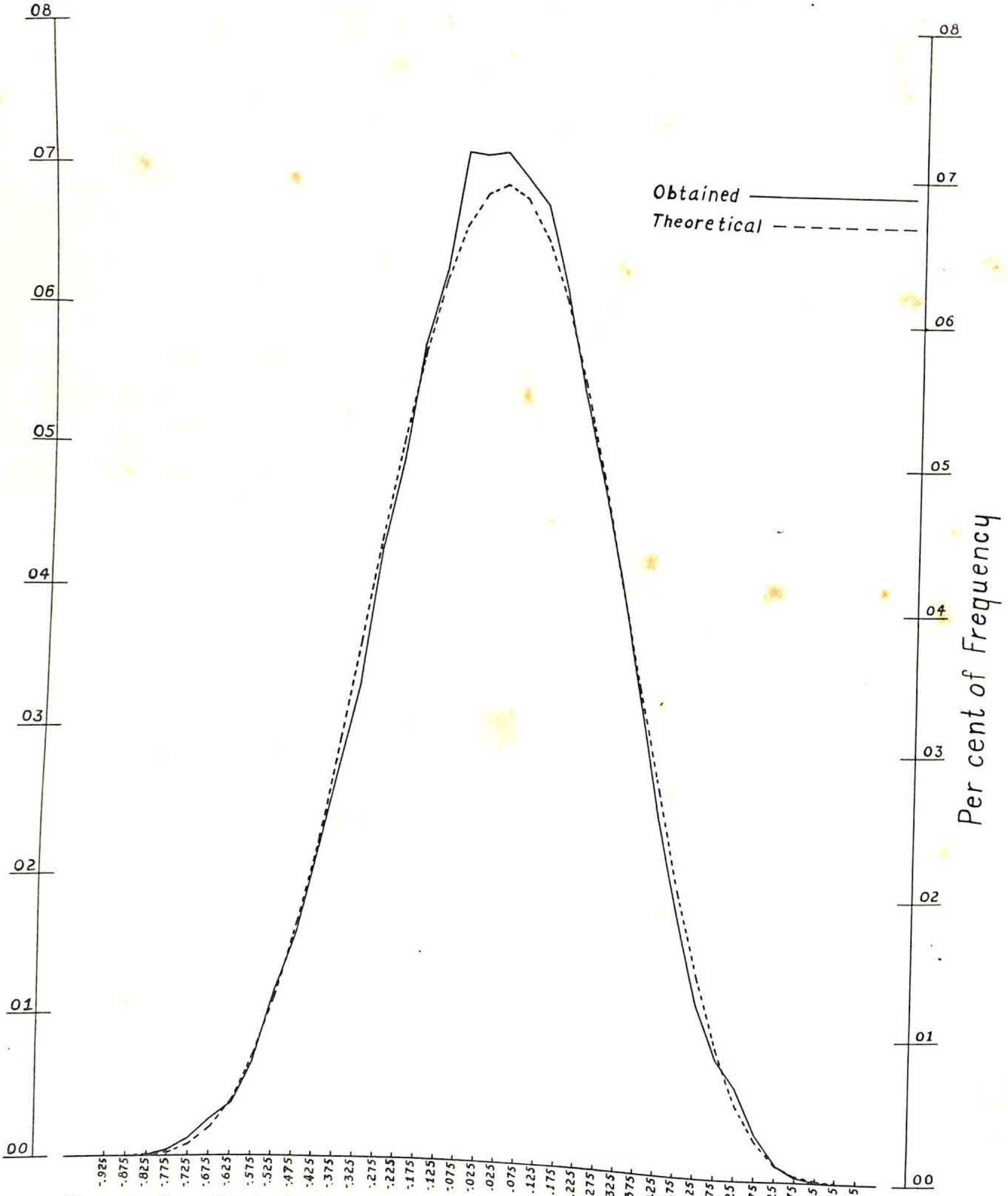
SIZE AND NUMBER OF SAMPLES FOR EACH EMPIRICAL SAMPLING DISTRIBUTION

Marginal Distribution	Population Correlation		Number of Samples		
	Small	Large	N = 15	N = 30	N = 90
Normal	.000	.835	15,984	7,992	2,664
Rectangular	.000	.831	15,984	7,992	2,664
Leptokurtic	.000	.809	15,984	7,992	2,664
Slightly Skewed	-.005	.826	15,984	7,992	2,664
Markedly Skewed	-.004	.826	15,984	7,992	2,664

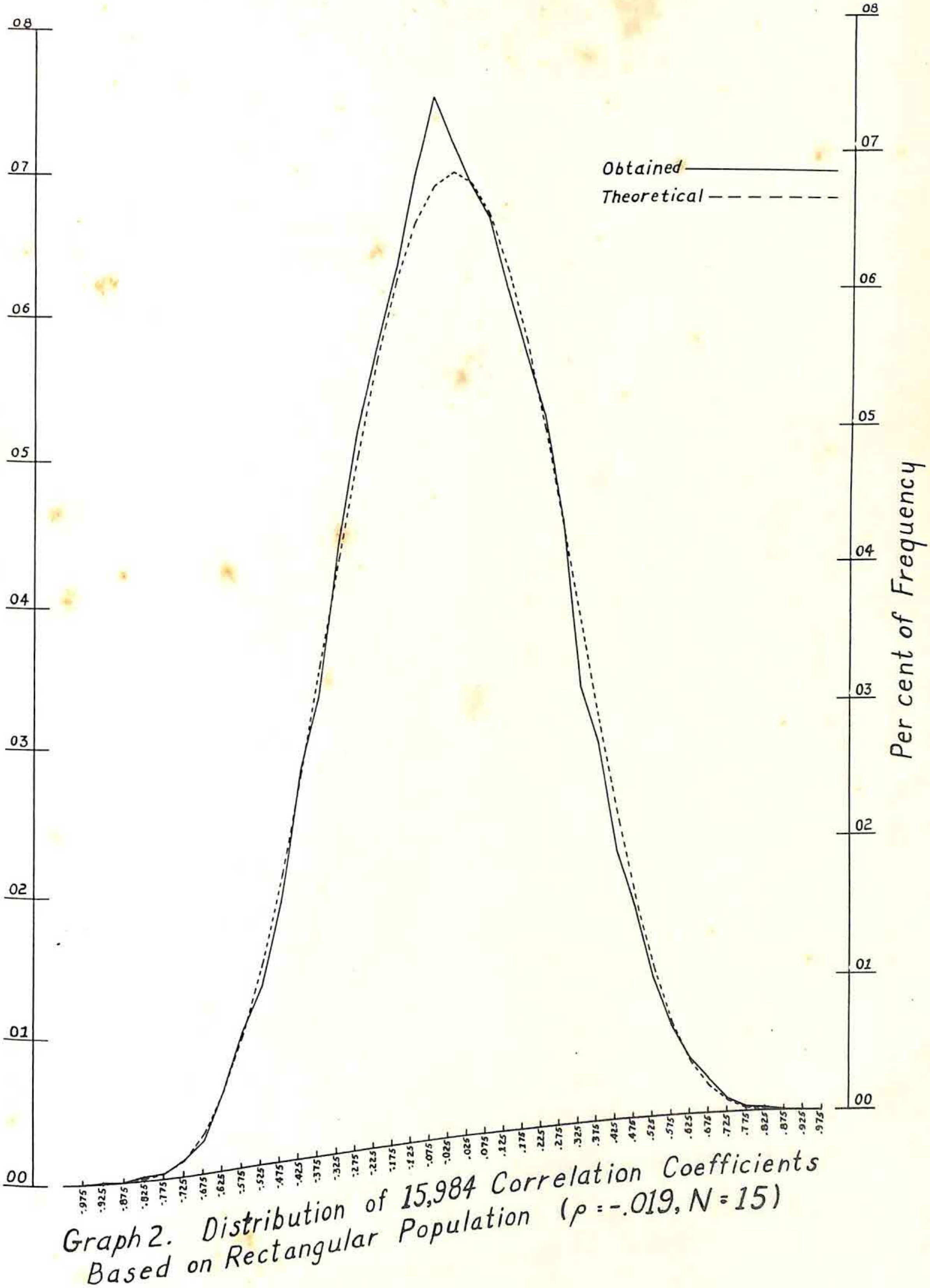
TABLE II

PROPORTION OF OBTAINED COEFFICIENTS IN CRITICAL REGION ($\rho \doteq .00$)

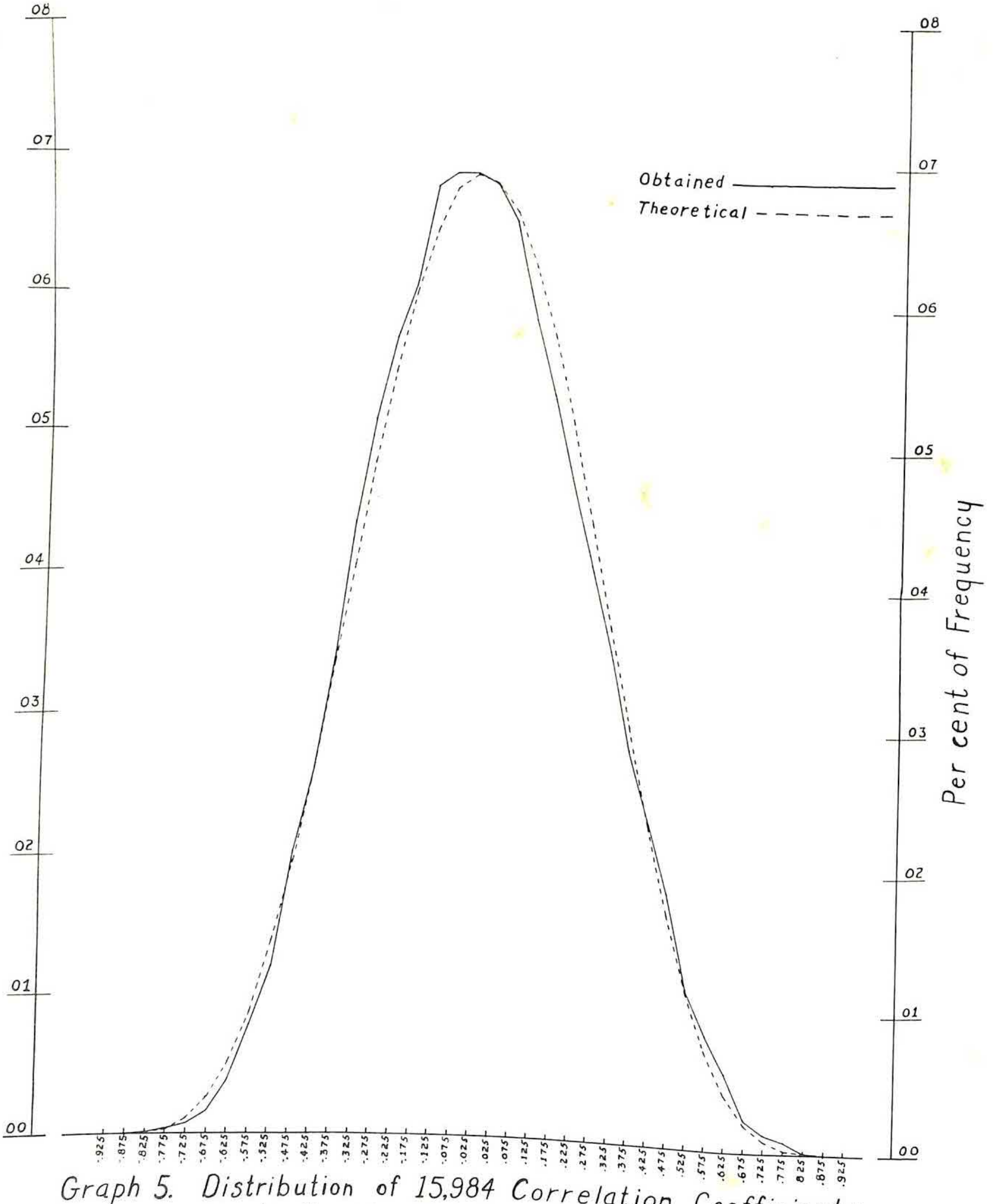
Marginal Distribution	Population Correlation	N	.05 Level of Signif.			.01 Level of Signif.		
			Lower Tail	Upper Tail	Both Tails	Lower Tail	Upper Tail	Both Tails
Normal	+.016	15	.0255	.0258	.0513	.0051	.0059	.0110
Rectangular	-.019	15	.0248	.0260	.0507	.0048	.0065	.0113
Leptokurtic	-.004	15	.0244	.0227	.0471	.0053	.0053	.0106
Slightly Skewed	-.010	15	.0220	.0256	.0476	.0052	.0063	.0115
Markedly Skewed	-.012	15	.0208	.0299	.0507	.0036	.0078	.0114
Normal	+.016	30	.0218	.0255	.0473	.0053	.0054	.0106
Rectangular	-.019	30	.0226	.0228	.0454	.0058	.0048	.0105
Leptokurtic	-.004	30	.0219	.0223	.0442	.0051	.0043	.0094
Slightly Skewed	-.010	30	.0224	.0274	.0498	.0046	.0068	.0114
Markedly Skewed	-.012	30	.0213	.0303	.0516	.0031	.0071	.0103
Normal	+.016	90	.0259	.0229	.0488	.0060	.0056	.0116
Rectangular	-.019	90	.0252	.0225	.0477	.0034	.0034	.0068
Leptokurtic	-.004	90	.0188	.0158	.0345	.0030	.0030	.0060
Slightly Skewed	-.010	90	.0199	.0267	.0465	.0041	.0071	.0113
Markedly Skewed	-.012	90	.0236	.0248	.0484	.0038	.0068	.0105



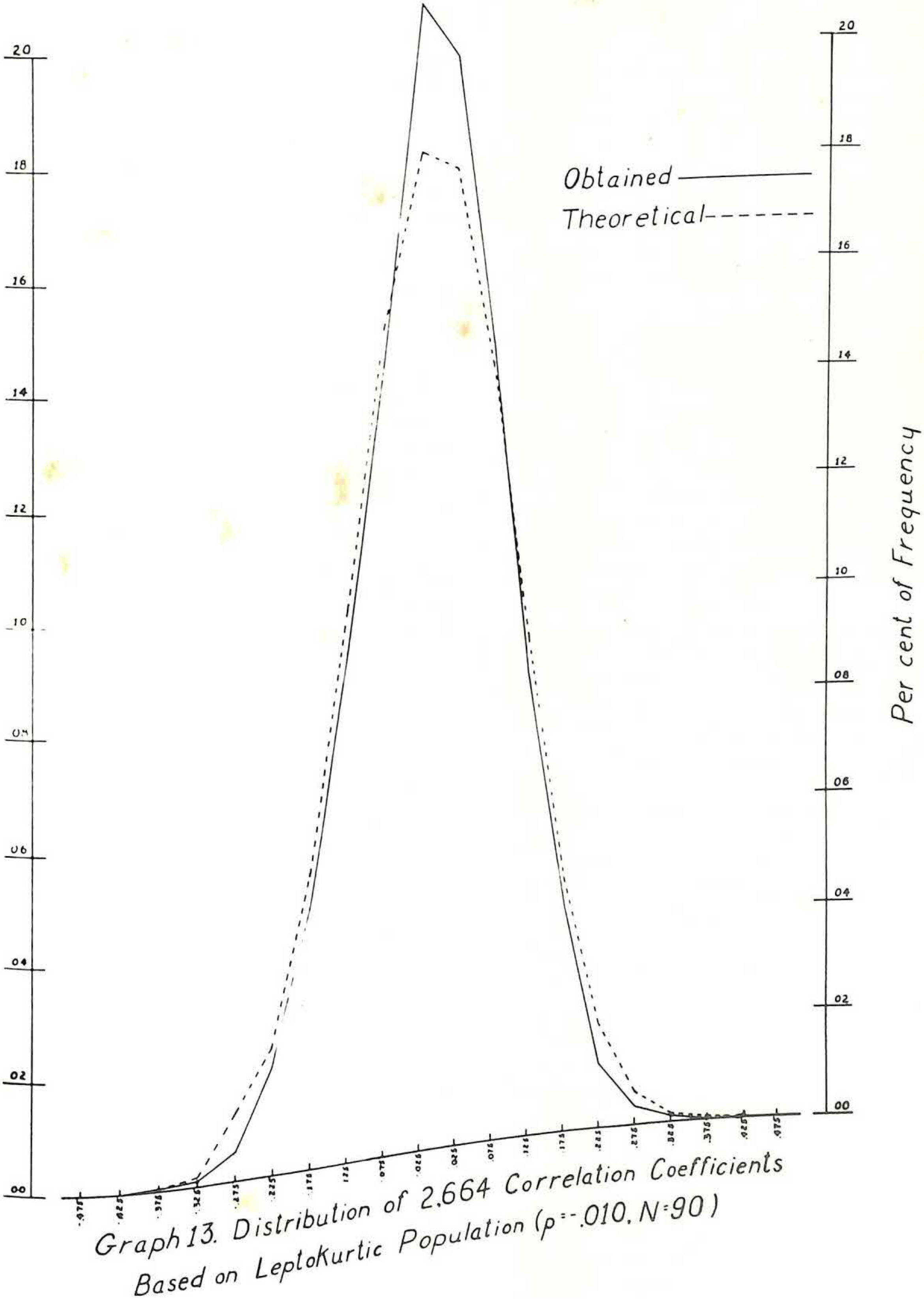
Graph 1. Distribution of 15,984 Correlation Coefficients Based on Normal Population ($\rho = +.016, N = 15$)



Graph 2. Distribution of 15,984 Correlation Coefficients Based on Rectangular Population ($\rho = -.019$, $N = 15$)



Graph 5. Distribution of 15,984 Correlation Coefficients Based on Markedly Skewed Population ($\rho = -.012$, $N = 15$)



Graph 13. Distribution of 2,664 Correlation Coefficients Based on LeptoKurtic Population ($\rho = -.010$, $N = 90$)

size 15 and 30 from a bivariate normal population having substantial correlation ($\rho \doteq .83$) deviated from their theoretical distributions at the .01 level of significance. For sample size 90 the Kolmorgorov-Smirnov test did not yield a significant difference at the .20 level. However, the deviation of the obtained and theoretical distribution in no case was extremely pronounced. Inspection of Table III shows that for both tails combined the incidence of significant coefficients approximated the theoretical number at the .05 and .01 levels of significance. However, there was a general tendency for there to be proportionately too many significant results in the positive tail of the distribution and proportionately too few in the negative tail. This disparity decreased as sample size increased.

The deviation of the obtained sampling distributions from the theoretical ones based on the assumption of normality of distribution for Fisher's statistic cast doubt on the validity of that assumption for values of rho as great as .835. For the purpose of determining the effect of nonnormality of population distributions on the sampling distributions of the product moment coefficient, the obtained versus theoretical comparisons seemed somewhat inappropriate. Consequently, the obtained sampling distributions based on nonnormal populations were compared also with the obtained sampling distributions based on the normal population.

The empirical sampling distributions for all sample sizes from the rectangular population having substantial correlation deviated from their theoretical distributions at the .01 level of significance. This departure was in the same direction as in the normal control population, but the deviation was much more pronounced. Table III shows that the critical regions of the obtained distributions in no case contained more than .477 of the number of significant coefficients obtained from the normal population. The disproportionately low incidence of significant coefficients became more extreme for increased sample size and for the lower level of significance. In one tail of the sampling distribution based on samples of size 90, there were no coefficients found as low as the theoretical value of $r_{.005}$.

All available evidence indicated that the sampling distributions of product moment coefficients based on a rectangular population were much more leptokurtic than either their theoretical distributions or the obtained distributions based on a normal population. At the .05 level of significance for both tails combined, the empirical distributions contained between .381 and .214 of the number of significant correlations found in the empirical distributions based on the normal population. At the .01 level of significance the corresponding proportions were .254 and .079. Rectangularity appeared to have a pronounced effect on the sampling distributions of the product moment correlation coefficient.

There was strong evidence that the sampling distributions based on a leptokurtic population containing substantial correlation were substantially more platykurtic than either their theoretical distributions or the obtained sampling distributions based on a normal population. Table III shows the proportions of significant coefficients in all critical regions were higher than anticipated from the theoretical distributions and from two to more than seven times as great as those obtained from the normal population.

All sampling distributions based on skewed populations having substantial correlation deviated from their theoretical distributions at the .01 level of significance. Table III indicates that all critical regions at the .05 and .01 levels of significance contained a greater proportion of significant correlations than expected from the theoretical distributions and a still greater proportion than found in the empirical distributions based on the normal populations.

Several general observations seemed warranted on the basis of the data from this study. They are:

1. When there was essentially no correlation in the population, the shape of the sampling distributions for the product moment correlation coefficients did not vary markedly as a function of nature or extent of nonnormality in the bivariate distribution. In general, these obtained sampling distributions conformed very closely to their theoretical distributions.
2. When there was substantial correlation in the population, the shape of the sampling distribution based on a bivariate normal population departed markedly from its theoretical distribution. Fisher's z transformation was apparently ineffective in providing a normally distributed statistic for values of rho as great as .835.
3. When there was substantial correlation in the population, the sampling distributions based on nonnormal populations deviated from their theoretical distributions and from the distribution empirically determined for a bivariate normal population. The nature and direction of the deviation was specific to the type of nonnormality in the population, some sampling distributions appearing leptokurtic and others being platykurtic.
4. Discrepancies between proportions of theoretical and obtained significant correlations were relatively greater at the .01 level of significance than at the .05 level. Some of the obtained sampling distributions contained too many correlation coefficients in the extreme tails and too few coefficients in the less extreme sections of the tails. In other sampling distributions the reverse was true. Due to the inconsistency of the effects of nonnormality on the tails of the sampling distri-

TABLE III

PROPORTION OF OBTAINED COEFFICIENTS IN CRITICAL REGION ($\rho \doteq .83$)

Marginal Distribution	Population Correlation	N	.05 Level of Signif.			.01 Level of Signif.		
			Lower Tail	Upper Tail	Both Tails	Lower Tail	Upper Tail	Both Tails
Normal	+.835	15	.0194	.0300	.0494	.0038	.0076	.0114
Rectangular	+.831	15	.0054	.0134	.0188	.0006	.0023	.0029
Leptokurtic	+.809	15	.0389	.0681	.1070	.0116	.0246	.0362
Slightly Skewed	+.826	15	.0255	.0405	.0661	.0065	.0112	.0177
Markedly Skewed	+.826	15	.0297	.0515	.0811	.0077	.0143	.0220
Normal	+.835	30	.0211	.0288	.0499	.0043	.0065	.0108
Rectangular	+.831	30	.0044	.0085	.0129	.0006	.0009	.0015
Leptokurtic	+.809	30	.0533	.0763	.1296	.0185	.0285	.0470
Slightly Skewed	+.826	30	.0310	.0439	.0749	.0123	.0116	.0239
Markedly Skewed	+.826	30	.0333	.0488	.0821	.0085	.0110	.0195
Normal	+.835	90	.0221	.0252	.0473	.0053	.0049	.0101
Rectangular	+.831	90	.0038	.0064	.0101	.0000	.0008	.0008
Leptokurtic	+.809	90	.0732	.0830	.1562	.0263	.0357	.0619
Slightly Skewed	+.826	90	.0312	.0417	.0728	.0131	.0116	.0248
Markedly Skewed	+.826	90	.0375	.0473	.0848	.0068	.0139	.0206

bution, it did not appear possible to adjust for nonnormality by modifying the level of significance in either direction.

5. When there was substantial correlation in the populations, increased sample size appeared to accentuate the discrepancy between the obtained sampling distributions and either their theoretical distributions or the empirical distribution on a bivariate normal population.

6. Sampling distributions based on skewed populations yielded disproportionate numbers of significant correlations in the two tails of the distribution whether there was correlation in the populations or not. Where rho was approximately zero, the two-tailed tests yielded approximately the expected proportion of extreme coefficients, but one-tailed tests based on the same distribution did not.

REFERENCES

1. Baker, George A. "The Significance of the Product-Moment Coefficient of Correlation with Special Reference to the Character of the Marginal Distributions," Journal of the American Statistical Association, XXV (1930), pp. 387-96.
2. Binder, Arnold, "Consideration of the Place of Assumptions in Correlational Analysis," American Psychologist, IV (1958), pp. 504-510.
3. Chesire, Leone, and others. "Further Experiments on the Sampling Distribution of the Correlation Coefficient," Journal of American Statistical Association, XXVII (1932), pp. 121-28.
4. Dunlap, Hilda Frost. "An Empirical Determination of Means, Standard Deviations and Correlation Coefficients drawn from Rectangular Populations," Annals of Mathematical Statistics, II (1931), pp. 66-81.
5. Furfey, P. H. "Comment on 'The Needless Assumption of Normality in Pearson's r,'" American Psychologist, XIII (1958), pp. 545-46.
6. LaForge, R. "Comment on 'The Needless Assumption of Normality in Pearson's r'," American Psychologist, XIII (1958), p. 546.
7. Milholland, J. E. "Comment on 'The Needless Assumption of Normality in Pearson's r;'" American Psychologist, XIII (1958), pp. 544-46.
8. Nefzger, M. D., and Drasgow, J. "The Needless Assumption of Normality in Pearson's r," American Psychologist, XII (1957), pp. 623-25.
9. Pearson, Egon S. "Some Notes on Sampling Tests with Two Variables," Biometrika, XXI (1929), pp. 337-60.
10. Pearson, Egon S. "The Test of Significance for the Correlation Coefficient," Journal of the American Statistical Association, XXVI (1931), pp. 128-34.
11. Pearson, Egon S. "The Test of Significance for the Correlation Coefficient: Some Further Results," Journal of the American Statistical Association, XXVII (1932), pp. 424-26.
12. Rider, Paul R. "On the Distribution of the Correlation Coefficient in Small Samples," Biometrika, XXIV (1932), pp. 382-403.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 3, March 1961)

PUPIL DISCOVERY VS. DIRECT INSTRUCTION*

WILLIS E. RAY
Ohio State University, Columbus

EMPIRICAL EVIDENCE does not support any one consistent set of hypotheses with regard to the relative efficacy of pupil discovery versus teacher dominated direct instruction (1, 2, 6, 14). This is particularly true as these teaching methods relate or interact with levels of intellectual ability of pupils (7, 8, 10).

It was the purpose of the present investigation (9) to provide additional experimental and applied research evidence as to the relative effect of directed discovery, in situations providing numerous problem solving opportunities, upon initial learning, retention, and transfer of micrometer measurement principles and skills as compared with traditional direct and detailed instruction in these situations, with three levels of intelligence.

The Problem

Many teachers instruct their pupils by the use of direct and detailed instruction. This method is sometimes characterized as the shortest road to the specific goal of efficient learning (3). By this method, the subject material being considered is presented to the pupils in great detail. The teacher poses problems and proceeds directly to the task of solving these problems without allowing much opportunity for the pupil to discover methods of solution or principles involved in their solution. The learner is actively or passively listening to and presumably assimilating principles and procedures for the correct solution of the problems presented. In many cases the pupils actually memorize the principles and procedures without fully understanding the relations among various items or steps which are presented one after the other by the teacher.

Two questions might be asked concerning direct and detailed instruction: 1) to what extent does the pupil retain ideas, facts, and principles learned in this way? and 2) how widely does the pupil apply this newly obtained material to new and related situations?

Studies in retention of learning reveal that active involvement of the learner in the learning situation retards forgetting (6, 11, 12). Research in transfer of learning has indicated that specific "rules" dealing with quantitative problems at least temporarily blind the learner so that he fails to adapt a learned approach to solve effectively novel problems requiring somewhat different techniques for solution (5, 13).

A method of teaching which would allow the pupil to become "active" in the learning situation may possibly have positive effects with regard to retention. If the pupil is allowed to discover relationships and methods of solution for himself, make his own generalizations and draw conclusions from them, he may then be better prepared to make wide applications of the material learned.

It is entirely possible that a directed discovery method of teaching may be most effective with pupils of high mental ability. Likewise, it is possible that a direct and detailed method of instruction may be most effective with pupils of low mental ability. The use of direct instruction, repetition, and drill have been common techniques used with the slow learner.

From these postulations the following hypotheses were generated and tested:

1. There is no difference in initial learning (knowledge of specific facts and principles, ability to solve problems, and actual manipulative performance) between groups taught by Method A and Method B.
2. There is no difference in retention of material initially learned as measured one and six weeks after instruction between groups taught by Method A and Method B.
3. There is no difference in the ability to transfer effectively the knowledge acquired as measured one and six weeks after instruction be-

---

* This report is based upon a dissertation submitted in partial fulfillment of the requirements of the Ed. D. at The University of Illinois.

tween groups taught by Method A and Method B.

4. There is no interaction between the two teaching methods employed and the low, average, and high intellectual levels with regard to initial learning, retention, and transfer of learning.

Procedure

Experimental Design. A treatments × levels design was employed in the experiment. Two experimental teaching methods (A and B) and one control (C) (no instruction) constituted the three treatments, and the three levels (high, average, and low) were determined with reference to mental ability (Pintner IQ) assumed to be related to the criterion. The levels were introduced to increase the precision of the design.

The independent variables were 1) methods of instruction and 2) intelligence. The dependent variables were 1) initial learning as measured immediately following treatment, 2) retention and transfer as measured one week after treatment, and 3) retention and transfer as measured six weeks after treatment. The controlled variables included 1) content and method of presentation under Method A and likewise under Method B, 2) length of instruction, 3) illustrations, 4) precision measuring equipment, and 5) testing and testing conditions.

The Sample. The 135 Ss initially selected for use in this experiment were sampled at random from all ninth grade boys enrolled in the three junior high schools of a medium sized (35,000 population) midwestern community. All 158 available Ss were assigned to the three levels by a "counting off" procedure suggested by Lindquist (4). By independent random sampling, 18 Ss were assigned to each of the experimental treatments within each of the three levels, and 9 Ss were assigned to the control group within each of the three levels. Fifteen Ss were lost due to absence during the six weeks of the experiment, and 3 Ss were eliminated at random to bring the number Ss in each experimental cell to 15. No Ss were lost in the control group. Therefore 117 Ss were considered in the descriptive analysis of attributes and in the analysis of the criterion data.

Due to the effects of random sampling, no statistically significant differences were found to exist within comparable cells with reference to intelligence, standardized achievement, and age. The Ss were quite representative of the general population of ninth grade pupils with regard to socio-economic status.

Learning Task. The zero to one inch micrometer caliper, a precision measuring instrument, was selected to be used in the teaching-learning process of the experiment. Ss were taught the names and functions of the parts, facts about the instrument, micrometer (readings to .001 inch) and vernier (readings to .0001 inch) principles involved in reading the tool, and how to manipulate and read the instrument for actual measurement.

Instruction. Ss were taught in groups of nine, each group being composed of three Ss from each of the three ability levels. The teaching-learning session was 47 minutes long. Common introductory material was presented to both treatment groups for the first 7 minutes, and the remaining 40 minutes were devoted to differential treatment. To assure constancy of conditions from one instructional group to another, tape recordings were used of all oral instruction and 35 mm slides were used of all illustrations (same set of 22 slides used with both groups).

The control group was given no instruction. Ss in this group merely took the five criterion tests to determine how they would perform without instruction.

Method A, as developed, represented a traditional approach to teaching skills and understandings, that of "direct and detailed" instruction, spoon feeding, or the "tell and do" method. By this method, the teacher presented the learning material, reviewed important points, and solved several examples related to the material without any pause other than normal sentence break throughout the forty minute session.

Method B, on the other hand, could be characterized as a method of "directed discovery". By this method the pupil was called upon to be active, to carefully study illustrative material on his own, and contemplate leading questions asked by the teacher. In Method B, 48 percent of the treatment time (19 minutes out of 40 minutes) was spent in silence which allowed the Ss to react to the leading questions, discover principles, and make generalizations. Very few positive statements of fact and none of principle or generalization were given orally by this method.

Criterion Measures. Three testing periods were used to evaluate the effectiveness of instruction. Immediately following treatment, the Initial Learning criterion test was administered (50 minutes). The Ss were given no indication that testing would occur one and six weeks after treatment. One week after instruction, the Retention One Week and Transfer One Week criterion tests were given (100 minutes). At a period six weeks after instruction the Retention Six Weeks and Transfer Six Weeks criterion tests were administered (100 minutes).

All tests were work limit or power tests and involved both performance and paper and pencil items. The initial learning and retention tests involved actual measurement of guage blocks with the micrometer caliper (performance) and paper and pencil items designed to test knowledge of facts and principles related to the use of the precision measuring instrument. The transfer tests involved both performance and paper and pencil

FIGURE I

RETENTION CURVES

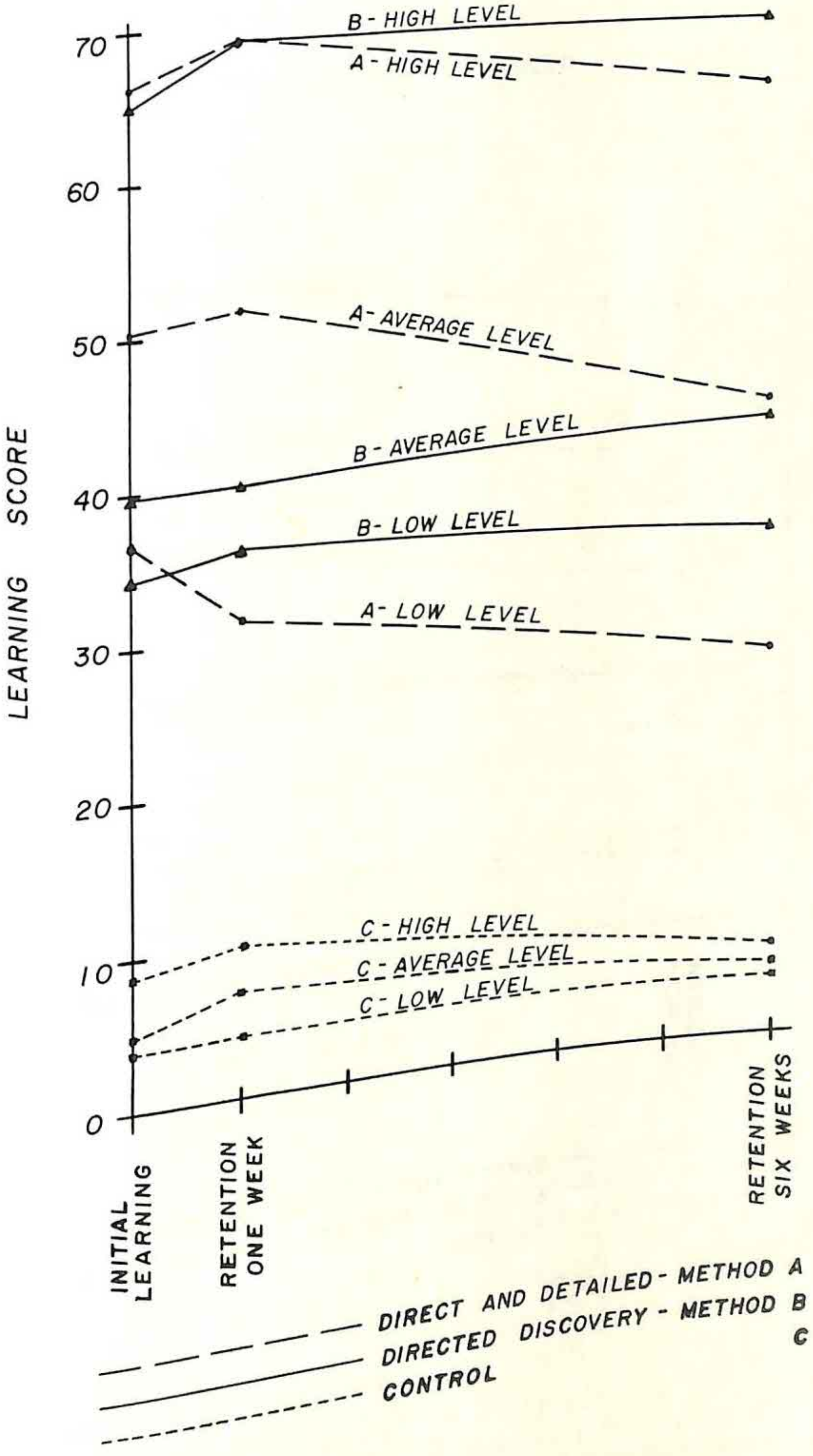
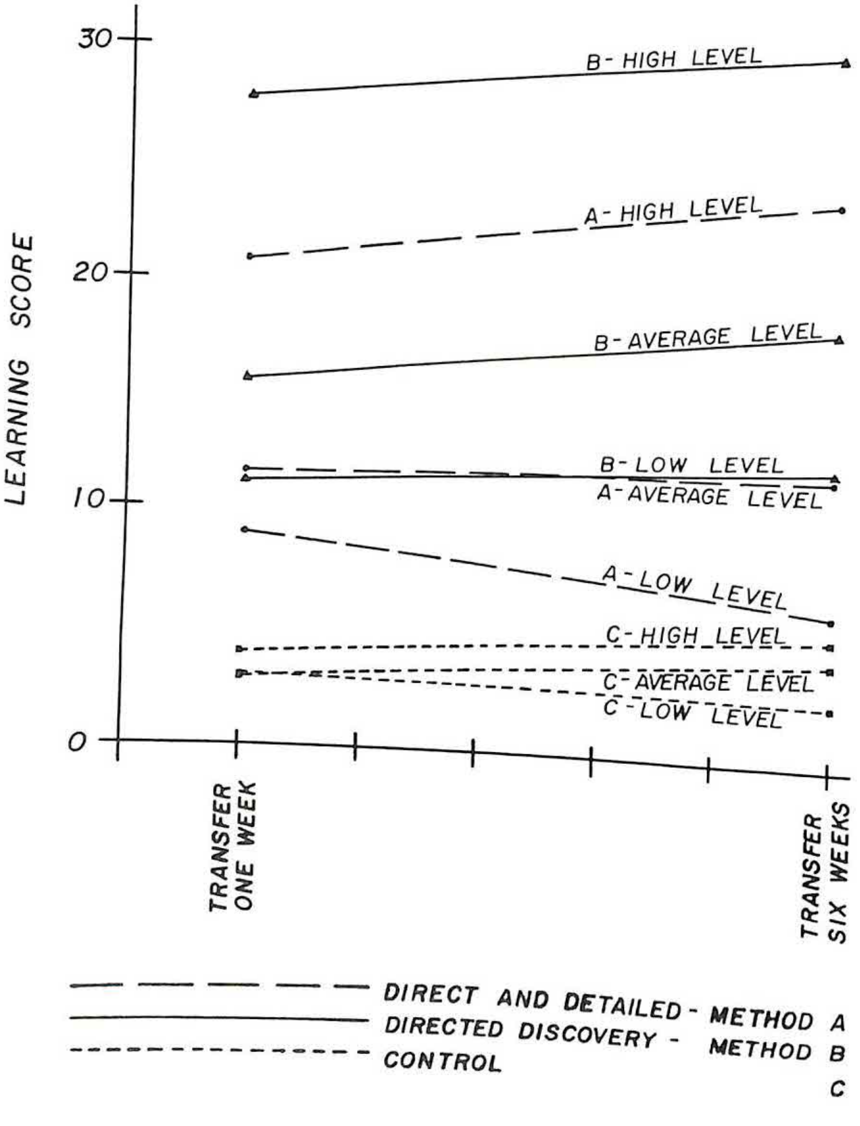


FIGURE 2

TRANSFER CURVES



items that attempted to measure application of knowledge gained to new but related situations. Performance with and knowledge of the depth micrometer caliper, the vernier caliper, and the vernier protractor (basic principles of use the same as with the vernier micrometer caliper) were measured by these transfer tests.

Subtests and part scores resulting from the tests were carefully analyzed by correlational techniques. As a result of this analysis, total test scores were determined by the sum of the raw scores of subtests.

Reliability (equivalence) coefficients ranging from .92 to .97 were obtained for the five criterion tests by the Kuder-Richardson Case IV (Formula 21) estimate.

Statistical Treatment. Data were punched on IBM cards and various electronic machines including the digital computer were employed in analyzing the data. The analysis of variance technique was used together with appropriate supplementary tests of homogeneity of variance. Correlational techniques and other methods employed were selected because of their adequacy in relation to the data. The .05 level of confidence and beyond was considered significant for the purposes of this study.

Findings

The control group was included in the design of the experiment to determine 1) if the ninth grade pupils used as Ss had prior knowledge of the learning task, and 2) if the Ss would obtain higher scores on tests subsequent to initial testing because of learning from the tests, vicarious instruction or communication among Ss between testings, or personal study. The performance of Ss in the control group as shown in Figures 1 and 2 indicates that none of the above factors was apparently operating in the experiment. Therefore, only the 90 experimental Ss were considered during the statistical analysis of criterion data.

Test of Hypothesis I. Hypothesis I--There is no difference in initial learning (knowledge of specific facts and principles, ability to solve problems, and actual manipulative performance) between groups receiving Method A and Method B.

An analysis of variance was performed on the obtained data, and this analysis is presented in Table I.

The analysis shows that the main treatments effect and interaction (T × L) account for a statistically insignificant amount of the variance associated with the learning scores of the groups. The F-test indicates that differences that are present between groups can be attributed to ability level. That is, significant differences in learning scores exist between levels only. The outcome with regard to levels was an anticipated one since the experiment was designed to develop real differences between levels. Resting on the data obtained and the statistical technique employed to analyze these data, the null hypothesis regarding initial learning is tenable.

Test of Hypothesis II. During the early design of this experiment, the investigator hypothesized that Ss receiving Method A would learn much more initially than those Ss receiving Method B. Of concern also, during the design stages of the experiment, was the fact that although Ss receiving Method A may learn more initially, they may not retain as much of what they initially learned as that retained by Ss receiving Method B. With this in mind, Hypothesis II was formulated as follows:

Hypothesis II--There is no difference in retention of material initially learned as measured one and six weeks after instruction between groups receiving Method A and Method B.

The clause "there is no difference in retention of material initially learned" or what they did in fact learn, was designed so that an examination would be made of the difference between the initial scores and scores made at one and six weeks. This technique of analysis would reveal whether Ss receiving Method A or Method B retained a greater proportion of what they initially learned.

To test this hypothesis, analyses of variance were performed using difference scores and the results are revealed in Tables II, III, and IV.

It can be seen from a study of Table II that there was no significant difference in difference scores between treatments one week after instruction. Since the available data offer insignificant evidence against the null hypothesis ($\bar{\Delta}I1_A = \bar{\Delta}I1_B$), this null hypothesis is thus tenable.

However, during the ensuing five weeks, the group taught by the direct and detailed method failed to retain as much of what they did learn when compared with the retention of material learned by Ss instructed by the directed discovery method.

The analysis that supports this statement follows. Since the F-value for treatments obtained from Table III is significant, the available data offer evidence against the hypothesis of no difference ($\bar{\Delta}16_A = \bar{\Delta}16_B$). Bartlett's test of homogeneity of variance produced a chi-square of 1.327 (table value at .05 = 3.841). This test gives support to the belief that the variances of the two treatments are homogeneous. Since the F-test applied to the same data resulted in the rejection of the null hypothesis, and since Bartlett's test indicated that it is not the variances that differ significantly, it appears that the significant value of F is the result of a difference in the means of the difference scores (D_{16}).

Considering the difference scores obtained by taking the difference of the scores attained one week after instruction and the scores six weeks after instruction, the evidence indicates that at

TABLE I

ANALYSIS OF VARIANCE OF LEARNING SCORES ON INITIAL LEARNING TEST

Source	Sum of Squares	df	Mean Square	F
Treatments	313.7	1	313.7	1.041
Levels	15,624.8	2	7,812.4	25.920*
Treatments × Levels	596.7	2	298.4	-
Within Sets	25,318.8	84	301.4	
Total	41,854.0	89		

* Significant F at .05 = 3.11 Significant F at .01 = 4.87

TABLE II

ANALYSIS OF VARIANCE OF DIFFERENCE SCORES BETWEEN INITIAL LEARNING AND RETENTION ONE WEEK

Source	Sum of Squares	df	Mean Square	F
Treatments	48.4	1	48.4	-
Levels	291.7	2	145.9	1.616
Treatments × Levels	365.2	2	182.6	2.022
Within Sets	7,584.3	84	90.3	
Total	8,289.6	89		

six weeks, the Ss taught by the directed discovery method (Method B), retained a relatively larger proportion of material that they had at their command one week after instruction when compared with Ss that received Method A.

The difference scores obtained from the initial Learning and Retention Six Weeks data were analyzed in Table IV. The F-value of 9.333 for treatments provides significant evidence for rejection of the null hypothesis of $A16_A = A16_B$. However, Bartlett's test of homogeneity of

TABLE III

ANALYSIS OF VARIANCE OF DIFFERENCE SCORES BETWEEN RETENTION ONE WEEK AND RETENTION SIX WEEKS

Source	Sum of Squares	df	Mean Square	F
Treatments	864.9	1	864.9	8.701*
Levels	30.4	2	15.2	-
Treatments × Levels	168.3	2	84.2	-
Within Sets	8,350.5	84	99.4	
Total	9,414.1	89		

* Significant F at .05 = 3.96

Significant F at .01 = 6.95

TABLE IV

ANALYSIS OF VARIANCE OF DIFFERENCE SCORES BETWEEN INITIAL LEARNING AND RETENTION SIX WEEKS

Source	Sum of Squares	df	Mean Square	F
Treatments	1,322.5	1	1,322.5	9.333*
Levels	479.4	2	239.7	1.692
Treatments × Levels	220.3	2	110.2	-
Within Sets	11,904.8	84	141.7	
Total	13,927.0	89		

* Significant F at .05 = 3.96

Significant F at .01 = 6.95

variance provided a significant chi-square of 5.258 (table value at .05 = 3.841 and at .01 = 6.635). Nevertheless, an F as large as 9.333 obtained from Table IV may be the result not only of heterogeneity of variance, but also the result of a significant difference in means.

A t-test can be performed even though the variances are found to be heterogeneous. The standard error of the difference in means is calculated by using one-half the degrees of freedom available (n-1 rather than $n_A + n_B - 2$). A t-value of 3.054 was obtained using this procedure (table value at .05 = 2.015 and at .01 = 2.691). Therefore, the null hypothesis of $\bar{\Delta}I6_A = \bar{\Delta}I6_B$ is not tenable.

Hence the $\bar{\Delta}I1_A = \bar{\Delta}I1_B$ portion of Hypothesis II is tenable. The $\bar{\Delta}I6_A = \bar{\Delta}I6_B$ portion of the research hypothesis must be rejected.

Test of Hypothesis III. Hypothesis III--There is no difference in the ability to transfer effectively the knowledge acquired as measured one and six weeks after instruction between groups receiving Method A and Method B.

Data were collected by means of the criterion transfer tests given one week and six weeks after instruction. The results of the analyses of variance are presented in Tables V and VI.

In both Tables V and VI the interaction component of the analysis accounts for an insignificant portion of the variance between groups. The variance associated with levels in both analyses is a large part of the total variance.

A significant difference in main treatments effect is present both at one and six weeks after instruction. At one week the F-value of 5.158 is significant beyond the .05 point. Bartlett's test of homogeneity of variance produced an insignificant chi-square of 1.003 (table value of chi-square at .05 = 3.841). Therefore, the variation between the variances of the two treatment groups is well within the limits of random sampling from a population with a common variance. Hence the null hypothesis of $\sigma^2_A = \sigma^2_B$ is tenable. It can be concluded from this that the difference is in the treatment means. The significant F provides support for the inference that Method B is superior to Method A with regard to a relatively wide application of learning as measured one week after instruction.

Six weeks after instruction the analysis of variance supplies an F of 10.885 which is significant beyond the .05 point. A chi-square of .678 obtained from Bartlett's test of homogeneity of variance is not significant (table value at .05 = 3.841). Therefore, since the null hypothesis of $\sigma^2_A = \sigma^2_B$ is tenable, the basis for the significant value of F must be the difference between the means of the treatment groups.

Based on the data presented and the statistical procedures used in analyzing these data, Hypothesis III must be rejected. The alternative hypothesis is that the mean scores of the treatment groups are not equal must be accepted. This leads to the conclusion that Method B is superior to Method A in producing transfer under the conditions of this experiment.

Test of Hypothesis IV. Hypothesis IV--There is no interaction between the two teaching methods employed and the low, average, and high intellectual levels with regard to initial learning, retention, and transfer of learning.

Data pertaining to this research hypothesis have been presented incidentally while testing Hypotheses I through III. In every analysis of variance performed, the treatments × levels or interaction component of the variance between groups has been insignificant. Therefore, in light of the data collected under the conditions of this experiment, the null hypothesis of zero interaction is tenable.

Discussion

It was thought prior to the experiment that the directed discovery method would be more effective with the brighter pupils and the direct and detailed method more effective with the less able pupils. The complete absence in this experiment of significant interaction between teaching method and mental ability was an unexpected result.

Of considerable interest is the shape of the curves depicting the retention of pupils taught by the directed discovery method. Even at a period six weeks after instruction, these curves do not indicate any appreciable loss of material learned.

The following finding has implications for the design of future research of this type. Had not the difference scores method been employed, significant differences in retention would not have been discovered. That is, by simply analyzing retention data at one week and again at six weeks, no difference in retention by treatment group was found. However, by comparing the difference between scores made by individuals initially and at one and six weeks, significant differences were ascertained.

It would appear that the directed discovery approach to teaching and learning can be used with the pupil of low mental ability with most satisfactory results. It has been held by many teachers that careful direct and detailed presentation with numerous examples, followed by repetition and drill, is an effective way of teaching the slow learner.

Teaching methods compared in this study are rarely used in pure form. Good teaching makes use of selected approaches geared to specific teaching situations. The results of this study are applicable to the extent that each method compared can be applied in practice in pure form.

These findings support some prior studies and are in conflict with results of other experiments previously noted. It would seem to the writer

TABLE V

ANALYSIS OF VARIANCE OF LEARNING SCORES ON TRANSFER ONE WEEK TEST

Source	Sum of Squares	df	Mean Square	F
Treatments	440.0	1	440.0	5.158*
Levels	3,362.5	2	1,681.3	19.710**
Treatments × Levels	86.1	2	43.1	-
Within Sets	7,164.3	84	85.3	
Total	11,052.9	89		

* Significant F at .05 = 3.96 Significant F at .01 = 6.95
** Significant F at .05 = 3.11 Significant F at .01 = 4.87

TABLE VI

ANALYSIS OF VARIANCE OF LEARNING SCORES ON TRANSFER SIX WEEKS TEST

Source	Sum of Squares	df	Mean Square	F
Treatments	877.3	1	877.3	10.885*
Levels	4,897.0	2	2,448.5	30.378**
Treatments × Levels	.6	2	.3	-
Within Sets	6,768.0	84	80.6	
Total	12,542.9	89		

* Significant F at .05 = 3.96 Significant F at .01 = 6.95
** Significant F at .05 = 3.11 Significant F at .01 = 4.87

that replication of experiments must be encouraged to build up a body of knowledge regarding theory of teaching and learning. In many instances, particularly in graduate research in education, a taboo is placed upon the replication of previous work.

Summary and Conclusions

In summary the investigator found:

1. With regard to initial learning, that there was no statistically significant difference in achievement between those Ss receiving Method A or Method B.
2. With reference to retention, at one week, of material initially learned, that there was no statistically significant difference in retention between Ss receiving Method A or Method B.
3. With respect to retention of material initially learned as determined six weeks after instruction, that Ss taught by the directed discovery method retained a statistically significant greater proportion of this learning when compared to Ss instructed by the direct and detailed method.
4. In regard to effective application or transfer of learning as determined one week after instruction, that there was a statistically significant difference in transfer, between Ss receiving Method A and Method B, in favor of the directed discovery group.
5. In relation to transfer six weeks after instruction, that Ss taught by the directed discovery method were more able to apply, to a statistically significant degree, the learning acquired when compared to application made by Ss given the direct and detailed treatment.
6. In attempting to discover differential effectiveness of direct and detailed and directed discovery teaching methods with the low, average, and high intellectual levels, that there was no apparent interaction between teaching method and intellectual level.

The following conclusions seem warranted:

1. The direct and detailed and the directed discovery methods of teaching are equally effective with regard to initial learning of micrometer principles and skills.
2. The direct and detailed and directed discovery methods of teaching are equally effective with reference to retention of material initially learned as measured one week after instruction.
3. The directed discovery approach to teaching is superior to direct and detailed instruction with respect to retention of material intially learned as determined six weeks after instruction.
4. The directed discovery method of teaching is more effective than the direct and detailed approach in enabling pupils to made wide applications of material learned to new and related situations, both at one and six weeks after instruction.
5. There is no interaction of teaching method and intellectual level.

REFERENCES

1. Fowler, H. L. "Induction or Deduction? An Experimental Investigation in the Psychology of Teaching," Educational Research Series, No. 5, Melbourne University Press, 1931.
2. Hendrix, Gertrude. "A New Clue to Transfer of Training," The Elementary School Journal, 1947, 48, 197-208.
3. Katona, G. Organizing and Memorizing, New York: Columbia University Press, 1940.
4. Lindquist, E. F. Design and Analysis of Experiments in Psychology and Education, Boston: Houghton Mifflin, 1953.
5. Luchins, A. S. "Mechanization in Problem Solving," Psychological Monographs, 1912, 54, No. 6 (Whole No. 248).
6. McConnell, T. R. "Discovery Versus Authoritarian Identification in the Learning of Children," University of Iowa Studies in Education, 9, No. 5. Iowa City: University, 1934.
7. Michael, R. E. "The Relative Effectiveness of Two Methods of Teaching Certain Topics in Ninth Grade Algebra," The Mathematics Teacher, 1949, 42, 83-87.
8. Nichols, E. D. Comparison of Two Approaches to the Teaching of Plane Geometry, Unpublished Ph.D. Dissertation, University of Illinois, 1956.
9. Ray, W. E. An Experimental Comparison of Direct and Detailed and Directed Discovery Methods of Teaching Micrometer Principles and Skills, Unpublished Ed.D. Dissertation, University of Illinois, 1957.
10. Sobel, M. A. A Comparison of Two Methods of Teaching Certain Topics in Ninth Grade Algebra, Unpublished Ph.D. Dissertation, Teachers College, Columbia University, 1954. (Microfilm).
11. Swenson, Ester, Anderson, G. L., and Stacey, C. L. Learning Theory in School Situations, Minneapolis: University of Minnesota Press, 1949.
12. Thiele, C. L. The Contribution of Generalization to the Learning of the Addition Facts, New York: Columbia University Press, 1938.
13. Thorndike, R. L. "How Children Learn the Principles and Techniques of Problem-Solving," in N. B. Henry (Ed.), Learning and Instruction, the Forty-ninth Yearbook of the National Society for the Study of Education, Part I. Chicago: University of Chicago Press, 1950, pp. 192-216.
14. Winch, W. A. Inductive Versus Deductive Methods of Teaching: An Experimental Research. Baltimore: Warwick and York, 1913.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 3, March 1961)

ANALYSIS OF QUESTIONNAIRE RESPONSES BY CERTAIN PSYCHOLOGISTS*

ANDREW L. COMREY
University of California at Los Angeles

A PREVIOUS paper reported the results of a survey of psychologists listed in the 1951 APA directory as being males, employed at colleges or universities, and members of Phi Beta Kappa (1). The questions concerned work habits, personal characteristics, interests, preferences, and attitudes. About ninety percent of the questionnaires were returned. Each question was statistically related by means of the epsilon coefficient to several internal questionnaire criteria: interest in administrative work in psychology, industrial consulting, counseling and guidance, psychotherapy, research, teaching, and publication rate. The last criterion measure was determined as a ratio of the number of scientific articles and monographs published to the years of full-time professional experience.

The considerable number of significant relationships discovered suggested that further analysis might prove worthwhile. Accordingly, sixty of the most significant items were chosen to constitute the variables for a factor analysis. A random selection of 216 cases was taken for which phi coefficients were computed between each of the 60 items and each other one. Each phi coefficient was divided by the maximum phi possible for the marginal totals concerned. Since most of the items were multiple choice or completion items, it was necessary to dichotomize each one artificially. This was done as near the median as possible in each case.

Seventeen centroid factors were extracted from the matrix of phi over phi max coefficients. The two highest loadings on the last factor were .22 and .21, respectively. No loading of .3 or more occurred after the ninth factor. The 17 centroid factors were rotated analytically by Kaiser's Varimax method (2). This procedure tends to maximize the variance of the squared extended vector projections, which usually yields an orthogonal simple structure solution if one exists in the data.**

Results

Each factor will be presented with those variables for which the loadings were .3 or more. The factor number and proposed name will be followed by a figure in parentheses which gives the number of loadings between -.1 and .1 to indicate hyperplane density. After a brief statement about the factor, the item number and loadings will be given, opposite which will appear a description of the variable involved. Each variable is usually a dichotomized multiple choice or completion item, but only a question without the alternatives will be presented. It is understood that such questionnaire responses as "yes," "quite a bit," and so on, are toward the positive side of the variable, despite the fact that the precise alternatives will not be given here.

Factors I, VIII, and X were minor factors with only one or fewer loadings over .4; hence no attempt will be made to interpret them.

II. Administrative Interest (35). The two items of major importance on this factor clearly identify it as being concerned with the individual's desire to participate in administration.

Item	Loading	Description
45	.56	How much would you enjoy being head of the psychology department at a major university?
57	.55	Assuming you were qualified, how much would (do) you enjoy doing administrative work in psychology?
51	.31	How much do you enjoy working on group projects?

III. Teaching Interest (40). Item 57 was one of the internal criteria analyzed in the previous study and helped to define factor II. Item 62 also was such an internal criterion and represents an important defining variable for this factor.

Item	Loading	Description
36	.66	Would you object to a position, otherwise desirable, if you had no classes to teach?
62	.60	Assuming you were qualified, how much would (do) you enjoy

---

*All footnotes will be found at end of article.

TABLE I

PHI OVER PHI MAX COEFFICENTS

No.	1	3	5	7	8	10	12	14	16	18	20	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
1																														
3	79																													
5	-01	-05																												
7	04	06	25																											
8	19	17	-20	-11																										
10	00	-07	09	-01	-29																									
12	-02	-03	12	19	-16	14																								
14	-01	02	-09	-16	10	-05	-21																							
16	-11	-03	-05	08	15	-05	09	-07																						
18	-04	-05	00	13	05	-05	06	-04	39																					
20	37	44	12	28	-06	00	26	-14	07	11																				
22	-18	-12	03	-03	-04	05	00	00	18	11	-11																			
23	-08	-18	-13	-14	-04	-13	-05	-01	-04	12	-18	-02																		
24	-07	05	30	23	-29	07	27	-23	00	-05	32	-04	-05																	
25	-02	-02	21	20	-11	12	22	-19	-05	-15	38	-12	-17	40																
26	00	-12	-06	-06	12	03	05	08	00	-01	-05	-05	08	-10	-01															
27	-06	-09	14	27	-07	16	14	-23	03	04	19	-10	-04	09	18	-07														
28	19	20	-14	00	01	-09	-03	05	02	03	08	-05	20	-04	01	08	01													
29	20	14	-30	-12	13	01	-01	-04	-14	-07	-05	05	-06	-19	-16	13	00	06												
30	-11	-06	-04	07	04	02	02	-04	05	06	06	02	12	-01	04	03	13	05	03											
31	-02	-02	00	-11	13	-03	-17	02	04	-01	-03	01	37	-07	-13	01	-01	00	-09	-05										
32	-13	-06	00	01	02	05	09	-10	08	-05	01	-07	-22	03	26	-14	03	-22	-10	05	00									
33	-06	-12	-07	00	02	-09	00	-03	09	14	-06	00	19	-01	-02	-03	-09	-05	01	-13	14	-09								
34	-12	-14	03	09	-01	06	11	-13	03	06	13	-01	13	01	09	00	10	-15	05	20	-12	12	-12							
35	-32	-15	22	09	-08	10	10	-04	18	-01	17	-02	-01	24	25	-10	09	-13	-30	10	00	36	-15	13						
36	06	04	-07	-08	15	-32	-18	03	-07	-09	-14	01	-23	-08	-09	-09	-14	00	21	00	-18	05	-05	04	-02					
37	03	08	-15	-04	12	-23	-15	24	-06	-06	-31	19	-05	-22	-31	-10	-03	08	13	-06	08	-06	-04	-01	-09	18				
38	02	-02	-10	-23	06	10	-04	-02	-05	-02	-13	07	-02	-19	-14	11	00	-08	21	-06	15	23	-04	07	-05	13	04			
39	04	00	-09	-01	-02	-12	04	04	01	02	-06	06	-09	-01	-15	01	-10	00	05	-06	-01	-07	06	-11	-14	09	05	16		
40	00	02	02	05	05	-01	02	-08	09	04	11	11	-01	09	12	03	-07	-13	06	15	01	07	04	36	01	10	-21	09	22	
41	-16	-24	04	13	-17	-18	04	11	-05	04	-07	04	32	-05	-04	-01	00	-05	-18	14	03	-07	21	16	16	-15	-04	-17	-06	03
42	06	12	16	24	-03	-02	14	-23	07	00	39	-04	-09	29	26	05	16	-05	-22	03	-03	03	-03	04	41	-03	-16	-02	00	12
43	11	06	-09	-16	23	-21	-24	15	-03	-04	-22	-02	-02	-39	-11	00	-04	22	22	-13	02	00	-01	-18	-26	19	34	15	02	-21
44	12	08	20	13	-03	-07	-01	-04	-12	-01	-01	-08	-17	10	-01	-10	-05	-16	-02	-01	-01	09	-01	-09	14	09	14	01	14	-03
45	-01	07	05	-06	18	-12	-04	05	19	10	02	06	-13	-04	-02	-12	-08	01	06	10	04	11	-01	-02	10	09	-04	00	-04	10
46	-16	-19	-05	08	-08	08	12	-02	-04	-05	-10	03	07	-08	-14	-05	00	00	-02	15	-10	-01	10	10	-05	-02	08	05	05	06
47	-05	00	09	25	-08	01	28	-21	18	11	29	06	03	18	19	14	15	02	-12	13	00	09	05	06	18	-12	-30	-03	00	07
48	10	02	23	01	13	05	-06	11	03	-02	14	02	-09	-15	-04	-03	00	-01	08	04	07	15	-06	08	-09	14	03	06	-02	-02
50	-14	02	04	-01	00	01	-06	13	07	04	-08	08	07	09	01	-01	02	-07	05	12	-14	11	17	09	14	02	07	-09	-03	-02
51	-07	-01	05	-08	02	02	01	24	08	-01	00	05	05	-06	-01	-15	-03	02	-11	09	08	16	02	02	04	-04	15	-12	-07	-17
53	22	19	04	-03	-19	19	19	-07	-16	02	24	-22	07	08	16	-04	-05	04	01	15	-18	-12	14	11	00	00	-14	07	00	04
54	-08	-07	06	09	-12	-05	-10	13	-05	-06	-18	-04	-10	-02	08	-05	-06	-05	-03	-31	-01	11	-02	-28	-04	11	10	-02	07	-11
56	18	17	-14	06	00	-03	-10	-01	-07	-05	09	-13	15	-06	-11	10	02	14	05	-10	10	-15	12	02	-03	-01	-06	07	-06	-01
57	02	07	-14	10	05	12	01	15	08	-01	-01	11	13	-04	-11	-05	-02	17	-05	09	-01	-18	-05	00	07	-23	15	-10	-01	-04
58	-05	-09	-05	05	-23	16	14	05	-05	10	06	06	46	07	06	03	10	25	-06	07	-01	-12	13	01	-11	-44	00	-20	-13	-14
59	08	03	-23	-14	11	04	-07	25	-01	-02	-25	07	03	-33	-33	07	-14	25	12	-02	07	-11	-03	-04	-21	-01	21	18	07	-12
60	-10	-18	-11	-12	12	-09	-06	22	-01	05	-26	08	08	-09	-30	16	-06	-08	07	06	10	09	09	08	-12	10	10	24	23	05
61	-10	-01	31	19	-22	09	16	-21	07	03	23	-01	03	43	41	-01	24	02	-33	02	-03	15	05	-03	19	-29	-28	-10	-23	-03
62	04	00	04	-02	10	-01	-11	02	-02	01	-13	01	-10	-15	-03	-06	07	10	25	12	-07	05	-01	00	-07	17	25	13	-05	-05
63	-04	-03	16	02	-18	10	16	-20	-05	-14	16	-14	05	29	21	02	06	03	-23	-01	05	04	03	00	05	-09	-37	10	-08	00
65	01	06	-20	-17	03	15	-08	04	00	10	-01	08	12	-04	-09	-05	13	21	10	22	01	05	-07	05	04	07	10	06	-02	-08
66	-10	-11	-04	05	-03	16	10	-02	06	10	02	-06	28	06	04	07	10	01	-04	-02	24	-09	16	14	08	-22	-08	01	00	-01
67	01	02	-24	-20	09	01	-14	16	05	01	-13	09	09	-24	-17	00	03	20	16	06	-08	-09	-02	09	-04	08	11	05	-10	-11
68	17	24	-14	-06	-03	-07	-02	04	07	-06	15	08	-01	17	-04	03	-01	32	00	-03	-01	-16	-01	-23	-11	06	-02	-05	04	-05
69	24	21	-09	00	01	09	-04	14	-06	01	04	-04	15	-06	-10	10	-08	64	02	01	-12	-54	-11	-22	-14	-08	04	-28	-08	-08
70	20	21	-05	-01	05	-13	11	09	13	-12	13	04	04	-03	-05	09	01	25	06	-05	-03	-39	01	-18	-13	-05	11	-13	09	-04
71	23	18	10	24	06	-28	23	-19	-02	02	37	-12	-15	15	11	03	04	-01	-01	-06	-20	-21	-04	-07	-02	03	-06	-02	03	00
72	05	09	-11	02	-12	05	15	00	03	03	05	05	00	04	-07	04	05	44	-03	-03	-09	-23	-04	-23	-13	-15	05	-07	07	-10
73	02	-04	07	-07	07	-12	07	-06	-07	-02	03	00	11	-09	08	09	05	-03	-15	08	09	07	02	-05	14	-08	-11	10	-12	05
74	00	01	25	24	-16	13	26	-33	10	02	59	03	-15	47	49	-09	20	-12	-15	05	-10	06	-10	15	25	-19	-48	-14	02	12

TABLE I—Continued

No.	41	42	43	44	45	46	47	48	50	51	53	54	56	57	58	59	60	61	62	63	65	66	67	68	69	70	71	72	73	74
1																														
3																														
5																														
7																														
8																														
10																														
12																														
14																														
16																														
18																														
20																														
22																														
23																														
24																														
25																														
26																														
27																														
28																														
29																														
30																														
31																														
32																														
33																														
34																														
35																														
36																														
37																														
38																														
39																														
40																														
41																														
42	01																													
43	−10	−25																												
44	−07	07	06																											
45	−04	−05	16	11																										
46	34	02	03	−01	08																									
47	10	30	−24	−10	−01	17																								
48	00	−08	17	13	12	09	12																							
50	08	−12	02	16	21	03	−02	08																						
51	13	−09	09	12	−01	10	03	17	09																					
53	−07	08	−07	−10	−30	−13	12	−04	−16	01																				
54	−10	−01	17	05	−02	−12	−06	−04	00	01	−09																			
56	08	09	08	10	03	06	11	−10	06	00	02	03																		
57	28	−04	13	06	10	42	00	00	11	23	−04	01	11																	
58	27	−06	−03	−10	−10	14	05	00	08	05	05	−04	22	23																
59	03	−31	25	09	03	06	−27	12	−02	20	−06	−02	15	23	03															
60	01	−21	12	09	06	14	−04	26	04	13	−15	−02	05	02	08	34														
61	05	32	−20	07	−06	−05	30	03	03	−02	16	03	06	−17	02	−33	−21													
62	01	−18	40	14	13	11	−04	18	18	−04	08	03	08	06	−04	22	16	03												
63	12	31	−18	02	−07	−01	28	00	−04	−07	11	10	05	−17	08	−22	−16	39	−11											
65	10	−17	24	06	−03	16	12	16	01	22	13	−08	07	27	10	27	20	00	30	00										
66	23	−04	−02	−09	−04	02	−01	05	−02	−08	08	−01	14	03	26	02	04	01	−07	13	05									
67	−05	−37	21	−04	−05	13	−18	20	−02	14	06	−03	11	26	12	42	27	−20	35	−20	56	11								
68	−13	09	04	−08	00	−07	02	−05	−12	−04	00	03	20	00	06	04	−04	11	−03	08	09	−03	02							
69	08	−01	14	−17	−14	01	−06	−14	−14	−06	09	−02	13	27	27	15	−25	−09	08	−15	13	09	17	37						
70	−01	13	03	−06	00	06	07	−05	−02	00	18	−10	07	20	02	00	−17	02	−01	04	01	02	02	31	37					
71	−18	−02	13	07	01	−08	17	−07	−07	−03	09	19	06	03	−08	−14	−15	08	−19	06	−02	01	−12	20	01	42				
72	−04	04	06	−10	01	12	05	−13	05	−08	03	−08	22	26	13	14	−26	07	02	07	06	02	05	44	50	40	28			
73	−05	11	11	−01	04	05	10	06	02	00	08	−14	11	−04	00	01	−03	14	07	16	05	03	03	01	00	06	14	10		
74	−08	33	−33	06	−05	−06	31	07	−05	−10	16	−13	−07	−11	04	−33	−29	37	−19	30	−15	00	−16	05	−20	00	23	−01	13	

		doing teaching?
43	.36	How much would you like a well-paid teaching position in a small liberal arts college of good reputation in a small community?
58	-.31	Assuming you were qualified, how much would (do) you enjoy doing industrial consulting?

IV. Moralistic Orientation (35). These items embody a rigid adherence to established values, particularly in the religious and sexual area. It is noteworthy that in this sample "importance of religion as a force in the individual's life" should be so closely associated with intolerance of moral and philosophical deviates. Whether this would be true in a more representative sample of the general population would be difficult to say.

Item	Loading	Description
69	.75	How much would you resent a colleague who was an atheist?
72	.73	How much would you resent a homosexual?
28	.62	How much of a positive force is religion in your life?
68	.56	How much would you resent a a colleague who was an adulterer?
70	.46	How much would you resent a colleague who was a member of the communist party?
32	-.34	How would you describe your basic political philosophy (liberalism vs. conservatism)?

V. Drive (41). The exact nature of this factor is less clearly outlined than is the case with some of the others, but the hypothesized identity seems reasonable. The title is intended to mean something like energy level.

Item	Loading	Description
54	-.58	How much difficulty do you have in writing papers, speeches, etc. ?
34	.48	In comparison with other men, how much interest in sex have you had during your life?
30	.46	Are you at ease in social situations?
40	.37	Would you mind a job which required you to be at a particular place 40 hours per week?
43	-.33	How much would you like a well-paid teaching position in a small liberal arts college of good reputation in a small community?

VI. Seniority (38). These variables all fit well into the seniority category since each is related to length of service.

Item	Loading	Description
1	.85	Age.
3	.81	Years of full-time professional experience.
20	.54	About how many scientific articles and monographs have you published, or had accepted for publication?
53	.32	How much harder do you work than would be necessary to "get by"?

VII. Service Orientation (38). Psychologists are thought by the layman to be devoted to "helping people," but these results make it clear that variance exists among psychologists in the extent to which they are oriented in this particular direction. It is interesting that neither counseling or psychotherapy interest proved to be the dominant variable on any factor but rather these two joined with teaching interest as somewhat lesser variables on this factor of broader character. This would tend to suggest that considerations other than the desire to help people are also important in determining interest in counseling and psychotherapy.

Item	Loading	Description
67	.71	How much do you want to help people who need it?
65	.62	How much do you want to do something for society?
48	.44	To what extent would you allow graduate students' expressed interests to determine the content of a graduate course?
59	.41	Assuming you were qualified, how much would (do) you enjoy doing counseling and guidance?
62	.40	Assuming you were qualified, how much would (do) you enjoy doing teaching?
60	.35	Assuming you were qualified, how much would (do) you enjoy doing psychotherapy?
42	-.33	How much weight would you place on published research in evaluating the merit of a psychologist?

IX. Resentment of Plagiarism (44). Probably something a bit broader than the title would suggest is involved here, but it is difficult to say precisely what its nature is.

TABLE II
CENTROID LOADINGS

No.	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI	XVII
1	11	22	29	-51	08	29	-21	-25	23	23	-20	05	-07	-13	-13	08	-06
3	12	16	39	-49	08	15	-29	-18	31	18	-08	-07	-11	-05	-17	-07	06
5	20	-34	04	18	20	-13	-10	14	07	17	-13	16	-14	-06	09	17	-18
7	36	-17	16	05	07	-20	07	-12	-12	28	-14	-08	03	04	02	18	-17
8	-26	12	18	-13	-23	15	15	08	22	20	10	-10	16	-12	06	04	-11
10	16	05	-16	28	04	09	-31	-19	-17	-15	-12	03	03	05	-14	07	11
12	40	-11	14	10	05	-10	08	-16	-13	-18	-19	07	12	-11	-06	05	06
14	-32	28	-09	-04	08	-17	-16	-08	19	04	14	13	10	02	20	10	05
16	08	04	20	19	-28	-21	14	22	26	-05	-11	-26	08	04	-07	17	15
18	09	10	08	17	-27	-11	13	07	11	09	-22	-19	-06	09	-07	20	-05
20	57	-13	40	-13	-04	14	-20	-07	17	21	-14	04	18	-02	05	-05	12
22	-11	06	09	17	-16	-23	08	04	06	-18	-10	-07	-08	08	06	-02	15
23	15	44	-32	27	05	16	20	08	05	10	04	-08	-10	08	10	-17	08
24	46	-35	07	09	14	-13	-16	09	05	-06	-06	13	-16	20	-05	-12	10
25	40	-38	08	07	08	13	-26	15	-14	06	10	05	13	-05	-04	08	17
26	05	12	-08	-10	-05	16	19	-04	03	-15	06	07	18	11	06	16	-17
27	23	-08	14	15	04	10	02	06	-25	07	-12	-22	09	09	-06	-08	-15
28	16	50	13	-22	04	-06	-12	20	-11	-04	14	-05	04	16	-08	12	06
29	-28	14	13	-21	-18	22	10	-12	-21	02	-14	11	03	04	-16	-10	16
30	10	09	21	27	-23	05	-02	-10	-06	-03	24	-03	-10	-03	08	-08	-09
31	-08	12	-19	13	14	22	15	13	33	08	-09	-25	07	08	12	-13	04
32	-19	-44	17	30	15	14	-12	-05	03	-07	19	-24	20	-03	-22	08	20
33	02	14	-12	12	08	03	30	04	16	06	-13	15	-12	-10	-12	12	14
34	10	-10	13	30	-32	19	08	-30	-16	12	09	07	-03	10	13	-11	05
35	20	-31	10	30	06	-10	-10	05	05	09	26	-12	06	20	16	-10	08
36	-41	-18	29	-24	-10	07	12	13	-14	10	20	19	-26	10	11	11	12
37	-46	22	11	-11	17	-26	10	-09	-12	19	-04	22	-20	06	06	-14	06
38	-26	-04	13	-02	06	33	22	-14	-02	-26	-09	-09	11	13	-04	08	05
39	-12	-05	07	-13	05	-11	21	-18	08	-24	-18	09	-08	12	13	12	13
40	-11	-16	13	05	-30	11	19	-20	09	-04	09	19	-10	13	12	06	12
41	20	22	-22	43	12	-08	13	-18	-05	18	23	10	-11	-07	11	09	09
42	48	-29	14	-10	09	04	12	-06	20	05	22	-20	-06	16	12	08	04
43	-43	27	21	-16	21	05	09	23	-19	20	06	-03	13	-14	-07	09	02
44	-15	-17	24	06	30	-12	-03	-12	13	16	-03	12	-16	07	-09	-07	-22
45	-14	-02	25	12	-11	-16	10	13	20	14	13	10	15	-07	-24	-02	07
46	03	18	12	29	08	-11	22	-25	-20	-09	21	06	04	-21	07	03	05
47	42	-09	21	16	04	09	19	02	10	-13	06	12	07	-07	12	14	04
48	-18	02	29	22	07	17	-10	05	12	07	-11	18	13	-12	18	17	02
50	-07	03	11	24	02	-13	04	03	12	18	13	18	-06	06	-24	-14	06
51	-13	14	14	24	20	-14	-23	-07	14	02	07	-09	03	-18	14	-02	07
53	31	08	05	-11	07	34	-20	-11	-18	-10	-12	05	-29	-15	12	10	10
54	-19	-18	-12	-15	34	-18	-02	14	-12	13	-06	-06	14	14	04	19	10
56	13	25	06	-06	22	15	16	-04	04	11	09	11	16	27	-13	-12	-11
57	05	44	19	18	12	-35	-05	-21	-09	06	11	-06	13	-06	13	-08	09
58	31	43	-18	28	15	-04	-06	-05	-04	09	-07	04	10	10	-10	-08	06
59	-38	45	12	04	07	-03	-06	-17	-04	-16	-06	07	16	11	-02	04	-16
60	-41	11	09	26	08	07	14	-18	14	-17	-09	14	09	20	14	12	-10
61	47	-23	06	19	26	09	-08	30	10	-10	04	-08	-15	-04	-17	09	-13
62	-31	16	42	11	08	16	04	19	-32	12	07	17	-22	10	-15	22	-03
63	40	-17	-08	09	22	22	03	16	07	-19	13	12	-08	-10	-06	-06	02
65	-12	38	37	27	08	18	-19	11	-18	-16	04	-12	-09	06	12	-10	02
66	19	20	-18	22	10	15	17	07	-07	13	-22	10	15	17	08	-04	12
67	-28	47	24	20	-09	13	-25	13	-22	-10	-10	06	08	12	16	-09	-09
68	18	24	14	-33	08	-07	-04	23	11	-24	03	04	02	19	-03	-08	07
69	28	64	-06	-37	-05	-21	-23	15	-19	04	18	05	-04	18	09	22	-07
70	26	33	19	-36	09	-19	19	10	04	-09	-03	13	-03	-15	19	-11	08
71	28	-10	30	-37	10	-17	16	14	-09	04	-21	08	18	-26	18	-22	-13
72	29	37	16	-27	13	-34	11	16	-12	-28	11	07	09	06	-19	-11	-06
73	11	02	12	06	09	19	13	18	08	-06	16	03	08	-15	06	-06	-21
74	56	-40	20	10	-12	06	-16	11	05	-08	-17	12	14	-05	07	-10	08

TABLE III

ROTATED LOADINGS

No.	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI	XVII	h^2
1	01	-10	04	-10	02	85	-02	08	07	01	08	-02	08	-10	03	-10	-05	80
3	08	-11	01	-20	-02	81	00	-13	07	01	03	01	02	-09	-08	09	06	77
5	06	00	06	21	17	-02	-01	-05	04	-09	-41	-06	01	-05	-12	02	-40	44
7	-18	18	-04	07	09	12	13	04	06	-13	-19	-35	-13	-03	-04	13	-34	44
8	10	-07	17	12	-09	17	-06	20	21	26	25	04	-28	-02	-04	05	10	40
10	-13	10	-27	01	-03	-02	-14	-08	-41	-04	-17	-05	13	-01	12	-14	01	38
12	-23	22	-20	-04	-04	03	06	01	09	-25	-33	-12	-04	-11	05	-11	02	37
14	15	16	-05	-05	15	03	-17	07	-07	11	30	40	03	03	-05	10	-10	40
16	05	00	-04	-07	-04	-06	-01	-07	01	-08	-09	00	-67	04	-05	03	06	49
18	00	-04	-03	02	-09	00	00	00	-08	-12	05	-15	-48	07	00	-07	-15	32
20	-29	00	-15	-06	-12	54	-01	-04	19	05	-47	-07	-12	04	01	20	-07	74
22	02	02	-04	-05	-09	-19	-09	-17	-01	-23	10	11	-26	01	-03	02	05	23
23	12	08	-08	-11	-17	-12	-03	01	-06	11	10	-04	02	62	08	-23	-04	55
24	-11	-12	-10	-11	-02	-02	12	-22	-05	-20	-54	-03	10	05	-16	16	-15	54
25	-23	02	03	02	08	04	04	-11	-07	19	-61	-08	06	-07	08	15	02	53
26	-01	-03	-02	-09	-03	-03	02	45	02	-02	04	03	-01	03	11	-04	03	23
27	-12	-01	-07	01	-03	-03	-09	-01	04	-01	-16	-48	-02	05	03	06	01	30
28	-01	06	07	-62	06	13	-15	04	-08	12	08	-05	-08	05	07	-01	-03	49
29	-33	-13	15	00	-09	16	-08	00	01	-03	35	00	06	-07	00	-16	25	41
30	04	19	05	00	-46	-06	-13	-01	-04	06	-04	-09	-05	-02	-02	10	-02	30
31	25	-15	-12	14	09	01	06	-04	06	06	06	00	-10	48	01	-01	16	41
32	03	12	05	34	15	-07	-06	-07	-24	08	-28	-10	-08	-14	-17	23	47	65
33	02	08	10	05	06	02	11	04	00	-11	-02	12	-12	24	-10	-38	-02	29
34	-30	13	04	26	-48	-04	-04	03	-08	-02	-02	-08	02	10	01	14	-01	46
35	-01	11	-04	09	-07	-18	00	-11	-06	06	-33	-07	-04	15	-14	46	-03	45
36	-11	-13	66	05	-01	-02	-05	-03	07	-06	19	17	06	-17	-07	19	05	62
37	12	13	16	-01	18	01	-08	-29	08	-14	58	-09	03	07	-11	11	01	58
38	-04	-07	10	13	02	03	-13	26	-05	-25	12	-04	02	00	05	-01	44	40
39	-05	01	04	01	07	01	05	08	05	-46	11	22	-02	-03	04	03	07	30
40	-21	00	15	12	-37	02	13	14	-03	-18	-10	18	-09	03	00	12	00	35
41	03	50	00	05	-13	-15	06	00	-18	07	-05	04	07	36	-03	-12	-22	54
42	07	05	07	-07	-08	17	33	10	05	-11	-41	-11	-07	16	03	40	-01	56
43	00	10	36	-09	33	07	-27	02	17	22	33	-08	-02	-04	-08	-12	15	57
44	13	02	08	16	10	13	-07	-04	04	-18	-01	-09	16	-08	-44	10	-14	38
45	-05	06	09	01	-02	00	-01	-03	08	16	00	09	-30	-10	-41	-03	11	34
46	-02	56	03	00	-17	-16	-06	06	05	-07	03	-02	05	01	-03	-09	07	41
47	04	21	01	-03	-13	04	04	15	12	-12	-43	-11	-21	12	11	07	08	40
48	-03	10	12	24	06	13	-44	10	07	02	-11	16	-11	00	-07	-03	-01	37
50	-03	08	06	00	-09	-05	-01	-13	-07	08	-01	04	-05	09	-47	06	-02	28
51	23	31	-11	08	07	07	-28	-18	00	06	01	11	-03	02	-08	09	00	32
53	-03	03	11	-07	-15	32	-09	-09	-08	-12	-25	-05	26	01	40	-15	-06	49
54	-08	-01	10	02	58	-10	04	-02	-03	-01	05	00	05	01	00	16	-03	40
56	-09	01	-03	-24	01	14	-02	29	03	05	05	-14	15	27	-25	-01	05	25
57	-01	55	-19	-23	-04	02	-21	-13	06	01	21	-01	-06	09	-09	08	-07	53
58	-08	22	-31	-24	-02	00	-11	-01	-17	08	-04	-11	03	37	-06	-23	-13	48
59	06	12	-12	-13	04	03	-41	21	-09	-08	43	04	04	-09	-11	-09	07	51
60	06	06	00	21	01	-11	-35	32	-09	-28	22	18	-03	07	-18	01	09	50
61	24	-06	04	-10	04	-02	05	-03	-05	-03	-66	-26	-01	05	-07	-06	-07	59
62	-11	05	60	-08	01	01	-40	01	-13	-04	11	-17	01	-09	-19	-12	00	66
63	12	-12	02	-08	-07	-05	12	06	03	01	-55	-01	21	12	01	-14	08	43
65	11	12	11	-19	-19	00	-62	-12	-05	-03	-05	-15	04	10	04	02	15	57
66	-25	01	-12	-01	06	-07	-11	10	-01	-01	-06	-06	00	47	04	-19	-06	38
67	-06	-02	-01	-16	-15	-06	-71	-01	-04	07	26	-03	01	01	05	-05	00	64
68	04	-17	-02	-56	05	11	-03	04	13	-08	-06	11	-01	03	-01	05	08	41
69	-02	09	03	-75	04	12	-06	09	-11	18	19	02	00	02	26	01	-39	91
70	00	13	02	-46	-02	17	05	-03	47	-12	03	15	00	07	10	-08	08	54
71	-15	00	-06	-14	10	14	05	-03	70	-06	-14	-12	03	-16	05	03	-10	65
72	00	12	-10	-73	03	-06	05	04	21	-09	00	-08	02	-11	-12	-10	02	67
73	18	02	08	-02	-11	-03	-10	21	24	13	-20	-09	04	04	-04	-05	07	24
74	-27	-12	-21	05	-16	07	01	-10	17	-04	-67	-03	-08	-07	05	11	-05	68

Item	Loading	Description
71	.70	How much would you resent a colleague who was a plagiarist?
70	.47	How much would you resent a colleague who was a member of the communist party?
10	-.41	About how many evening hours, if any, do you spend on professional work in an average week?

XI. Research Productivity (27). Variable 74 which represented the ratio of number of publications to years of professional service, here was also one of the internal criteria used in the previous analysis. An interesting question which invites speculation is that of whether items 24 and 25 represent partial cause and effect relationships with item 74, and, if so, which is cause and which is effect?

Item	Loading	Description
74	.67	Production rate (ratio of articles to years of service).
61	.66	Assuming you were qualified, how much would (do) you enjoy doing research?
25	.61	How much of your professional career has been spent in positions where you were expected to do research?
37	-.58	Would you enjoy a job where 40 hours per week were devoted to teaching and preparation?
24	.54	Are you now employed by an institution which grants the Ph. D. in psychology?
20	.47	About how many scientific articles and monographs have you published, or had accepted for publication?
63	.55	How much do you want to have national recognition as a psychologist?
47	.43	How much influence do you expect your work to have on the development of psychology in the United States?
59	-.43	Assuming you were qualified, how much would (do) you enjoy doing counseling and guidance?
5	.41	To how many professional journals do you subscribe?
42	.41	How much weight would you place on published research in evaluating the merit of a psychologist?
29	-.35	Do most major universities place too much emphasis on research?
35	.33	How much emphasis should be placed on research methods in training clinical psychologists?
43	-.33	How much would you like a well-paid teaching position in a small liberal arts college of good reputation in a small community?
12	.33	About what annual income do you expect to reach?
14	-.30	About what percentage of psychologists measure up to your standards as to what a psychologist should be?

XII. Competitive Personnel Policy (37). The individual high on this factor is likely to feel that some psychologists are better than others and should be rewarded for it. He is also more active in attending professional gatherings.

Item	Loading	Description
27	.48	Should an outstanding man be promoted sooner than he otherwise would be if he received a better offer somewhere else?
14	-.40	About what percentage of psychologists measure up to your standards as to what a psychologist should be?
7	.35	How many regional and national psychology meetings did you attend in 1950, 1951, and 1952?

XIII. Entertaining (45). That item 18 should have a negative loading and item 16 a positive one seems reasonable in view of the fact that one cannot be two places at the same time. On the other hand, one would think that frequent home entertaining might call for frequent social ventures beyond the threshold too. Perhaps item 18 was interpreted by the respondents more in terms of public places of amusement.

Item	Loading	Description
16	.67	About how many evenings in an average month do you entertain friends in your home?
18	-.48	About how many evenings in an average month do you go away from home for an evening's entertainment?
45	.30	How much do you read for recreational purposes?

XIV. Materialistic Orientation (44). In one way or another, each item in this group has somethink to do with activities or events likely to enhance the financial position. Thus, the factor seems to be concerned with the extent to which great value is placed on money.

Item	Loading	Description
23	.62	Could you be interested in a business career if a really good opportunity came along?
31	.48	If you were to receive a large inheritance, would you give up your job?
66	.47	How much do you want to have quite a bit of money?
58	.37	Assuming you were qualified, how much would (do) you enjoy doing industrial consulting?
41	.36	Would you like to direct a research project involving a staff of 100 people or more?

XV. Hedonism (42). These items seem to characterize the highly verbal, fun-loving scholar who refuses to take himself or psychology very seriously. He feels that his duty is to enjoy life for himself and also for those other psychologists who are so busy or preoccupied that they haven't time to do so, or perhaps don't know how.

Item	Loading	Description
50	.47	How much time do you spend in "socializing" during working hours?
44	.44	How much do you like bibliographic work?
45	.41	How much do you read for recreational purposes?
53	-.40	How much harder do you work than would be necessary to "get by"?

XVI. Introversion (35). Two of these items place great importance on research and research training while the other finds being alone not unpleasant. The most obvious common denominator in these responses appears to be a type of self-contained, as opposed to people-oriented, personality.

Item	Loading	Description
35	.46	How much emphasis should be placed on research methods in training clinical psychologists?
42	.40	How much weight would you place on published research in evaluating the merit of a psychologist?
33	-.38	Do you find it unpleasant to be alone?

XVII. Collectivism (45). Many individuals in psychology emphasize harmonious merging into the team where each cooperative person is accepted and there is a minimum emphasis on personal evaluation. Self-seeking by energetic and ambitious group members is frowned upon as contrary to group goals. Cooperation replaces competition. The items loading on this factor appear to define such a general point of view.

Item	Loading	Description
32	.47	How would you describe your basic philosophy (liberal vs. conservative)?
38	.44	Do you believe that life in our society is too competitive?
5	-.40	To how many professional journals do you subscribe?
69	-.39	How much would you resent a colleague who was an atheist?
7	-.34	How many regional and national psychology meetings did you attend in 1950, 1951, 1952?

Discussion

Several of the criteria selected for analysis in the original published study emerged here as independent factors. Administration and teaching interest variables defined two factors individually while research interest and publication rate defined a third. Interest in industrial consulting failed to define a factor itself but had a substantial loading on Materialistic Orientation. The two remaining criteria, interest in psychotherapy and interest in counseling and guidance, showed up together on the Service Orientation factor, although with lower loadings than one might expect. Other variables not originally conceived as criterion measures proved to be important in factor definition.

With respect to the universe of discourse under examination, it would appear that the most significant independent dimensions along which these psychologists differ are: interest in administration, interest in teaching, moralistic orientation, seniority, desire to help others, research productivity, and materialistic orientation. It is interesting that each and every one of these factors represents an area frequently involved in personnel action and policy decisions. It is also interesting and noteworthy that these factors should have emerged from the application of an entirely analytic rotation criterion.

Summary

A factor analysis was carried out on 60 items from a questionnaire given to males listed in the 1951 APA directory as Phi Beta Kappa members. The questions concerned work habits, personal characteristics, interests, preferences, and attitudes. Seventeen centroid factors were extracted

from the matrix of phi over phi max coefficients. Analytic rotations were carried out using Kaiser's Varimax method. Factors of sufficient importance to warrant naming were: Administrative Interest, Teaching Interest, Moralistic Orientation, Drive, Seniority, Service Orientation, Resentment of Plagiarism, Dislike for Mathematics, Research Productivity, Competitive Personnel Policy, Entertaining, Materialistic Orientation, Hedonism, and Collectivism.

FOOTNOTES

* This study was made possible primarily because of the availability of SWAC, an electronic computer operated by Numerical Analysis Research of the University of California, Los Angeles, and supported by the Office of Naval Research. The opinions expressed are the author's. Further financial support for this work came from a research grant by the University of California.

**The tables of correlation coefficients, centroid loadings, and rotated loadings have been deposited with the American Documentation Institute. Order Document No. 4858 from the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for 35 mm. microfilm or for 6 × 8 in. photocopies readable without optical aid. Make check payable to Chief, Photoduplication Service, Library of Congress. A copy of the complete questionnaire together with the response frequencies also has been deposited.

REFERENCES

1. Comrey, Andrew L. "Publication Rate and Interests in Certain Psychologists," American Psychologist, XI (1956), pp. 314-22.
2. Kaiser, Henry F. "The Varimax Criterion for Analytic Rotation in Factor Analysis," Psychometrika, XIX (1954), pp. 173-82.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 3, March 1961)

APTITUDE AND COURSE ACHIEVEMENT IN A SHORT MISSILE TECHNICIAN COURSE*

HARRY E. ANDERSON, Jr., and FRANK GUENTER**
USAADHRU, HumRRO

Introduction

THIS STUDY seeks to determine limited aptitude characteristics of students in a short missile technician's course, and the general relationship of these aptitude characteristics to various parts of the course and the final course average. The Personnel Research Branch (PRB) of the Adjutant General's Office, Department of the Army, has put forth much effort to stabilize and standardize measures of aptitude characteristics. HumRRO has been concerned with training problems and the assessment and improvement of training circumstances. The purpose of the present study is to assess academic achievement of missile technicians in terms of their aptitude characteristics, and suggest concrete referents for further training research in this area.

The ten aptitude tests in the present Army Classification Battery (ACB), which are administered to most enlisted personnel when they enter service, became operational in 1949 (10). Since that time, PRB has continually reviewed and analyzed these measures (e. g., 2, 3, 10, 11) to ensure their stability and enhance their usefulness as predictive measures of success in various army occupational specialities. Though research is presently programmed for the ACB in the missile field, such information is not now available.

One major interest in HumRRO's training research has resided in comparing course achievement with later proficiency in the field. An exemplary study (9) determined the major operational functions for missile maintenance personnel, built proficiency and written tests to measure ability regarding these functions, and compared recent school graduates with experienced field personnel on the written and proficiency tests as well as several other variables. The aptitudes of the two major groups (i. e., the recent graduates and the field-experienced personnel) were not compared with regard to most of their ACB scores. Aptitude measures in this study were comprised of an electronic score (i. e., a compound of two ACB scores) and the Electronic Placement Test[1]; aptitude, as measured by these variables, correlated to some extent with written test scores but demonstrated very little relationship with the proficiency test scores. Now, it is not reasonable that aptitude should be unrelated to proficiency (i. e., ability); therefore, a more extensive study of aptitude characteristics of such groups might well prove rewarding from both the situational and measurement standpoint.

Another HumRRO study (1) investigated the relationship of academic achievement in various parts of a course. But again, no extensive use was made of aptitude as related to the various parts of the course.

Method

The Course—The course under investigation was a five and one-half week technician course conducted at the U. S. Army Air Defense School, Ft. Bliss, Texas. The 220 hours in the course were used as follows: 1) 21 hours for Missile Mechanic (MM) training; 2) 34 hours for Launcher Area (LA) training; 3) 85 hours for Missile Preparation and Depreparation (MPD) training; 4) 80 hours for non-academic functions, distributed variously throughout the course (e. g., incoming and outgoing processing time, physical training, etc.). The course is, essentially, a transition-type course following a much longer technician course.

A written examination was given after each of the three main parts (i. e., in this order, MM, LA, and MPD) of the course. A total course average (TCA) is computed by weighting the first two parts of the course one each, and assigning a weight of two to the last, longer part of the course. There were four classes, including a total of 119 enlisted trainees, for whom complete data were available.

The ACB Tests—Eight of the ten ACB test scores were available for the 119 trainees as follows:

Reading and Vocabulary	(RV)
Arithmetic Reasoning	(AR)

---

*All footnotes will be found at end of article.

Pattern Analysis	(PA)
Mechanical Aptitude	(MA)
Army Clerical Speed	(ACS)
Shop Mechanics	(SM)
Automotive Information	(AI)
Electronic Information	(EI)

The remaining two ACB tests (i. e., Army Radio Code Aptitude and Radio Information) were not included in the analysis because of scoring difficulties and lack of data. A complete description of these variables can be found in a previous publication (3) and in various other PRB studies.

The Analysis—The eight ACB test scores and the four course scores, including the TCA, were intercorrelated for each class separately, and the correlations combined by a weighted average using Fisher's z-transformations (e. g., 6:325-6). The resulting correlations between the ACB scores were factored by the centroid method, and the centroid factors were rotated graphically. The four course scores were then projected onto the rotated factor structure.

Results

The intercorrelations among the eight variables are presented in the lower half of the central portion of Table I, while the upper half contains the residuals after the extraction of two factors. The right-hand portion of Table I presents the correlations of each of the four course scores with each of the eight ACB variables. The means and standard deviations for the eight ACB variables are presented also appropriately to the left of Table I.

The Factor Analysis—The 8 × 8 intercorrelation matrix in Table I was factored by the centroid method, the communalities being estimated as the highest correlation in each column, respectively. Actually, four factors were extracted in a preliminary analysis, but graphic rotations on the Zimmerman apparatus failed to produce anything on the third and fourth factors. Consequently, only the first two centroids were used in the final analysis; these factors are presented in Table II. The reader will note that the first centroid has projected variances equal to 3. 4420; the second centroid, . 7742. The total variance, using communalities in the diagonals, is equal to 4. 8000. The two centroids, therefore, account for approximately 85 percent of the variance in the system.

The two centroid vectors were rotated graphically through a clockwise angle of $41^{o}\ 30'$. The rotated loadings are presented in Table III.

The first factor is a Technical factor determined, for the most part, by MA (. 718), SM (. 751), AI (. 614), and EI (. 768); RV and PA also have fairly high loadings on this factor, suggesting some reading and spatial characteristics in the Technical factor. The second factor is an Academic factor with RV (. 596), AR (. 753), PA (. 612), and ACS (. 513); MA and SM also have appreciable loadings on this factor, indicating some technical aspects of the factor.

The ACB and the Course Tests—The four course tests were projected onto the rotated factor structure (5:209-19), and the resulting loadings are presented in Table IV. The second test, LA, and the final course average, TCA, have greater loadings on the Technical factor than on the Academic factor. The reverse is true for the first test, MN; it has more relation to the Academic factor than to the Technical factor. The third test in the course, MPD, which represents the greater amount of instructional time and is given the most weight in the computation of the final course average, does not seem to have much in common with the system, but most of the attending communality is located on the Technical factor. Several inferences are available from this analysis, as will be presented in the Discussion section.

Discussion

The ACB Scores—The results of the factor analysis indicate that aptitude characteristics of missile technician trainees with respect to the ACB tests are of two major, distinct types, viz., academic and technical. This outcome supports the methodology used several times in previous HumRRO research (9), the use of both written and practical tests as measures of proficiency. The use of both types of tests may be required to tap both types of aptitudes and abilities of which the technicians show definite capabilities. The result also lends support to the U. S. Army Air Defense School's policy of using written and practical examinations in most of their courses for technicians; logically it would seem that, if both types of examinations are given, success on practical examinations will be related to technical aptitudes while success on the written examinations will depend to some extent on both types of aptitudes. Hence, a student's exhibition of content absorption throughout the technical courses in the U. S. Army Air Defense School will depend on his technical aptitudes and, to a more limited extent, on his academic-type aptitudes. This is not to say that a given course should be built towards aptitude characteristics of the students, but only that the measurement of learning throughout any course should take into account the goals of the course, the nature of the program, and the post-graduate demands that will be placed on the trainees. Further studies regarding aptitude characteristics of successful field-experienced technicians should indicate the propriety of measurement characteristics in the School's testing program.

The above discussion and the dichotomy of ap-

TABLE I

INTERCORRELATIONS, RESIDUALS, MEANS, AND STANDARD DEVIATIONS OF THE ARMY CLASSIFICATION BATTERY TESTS, AND THE CORRELATIONS OF THESE TESTS WITH THE COURSE VARIABLES*

Means	Standard Deviations	Tests									Course Variables			
			RV	AR	PA	MA	ACS	SM	AI	EI	MM	LA	MPD	TCA
119.73	17.30	RV		.063	-.113	-.017	-.040	-.017	.035	-.026	.460	.401	.118	.404
111.87	15.35	AR	.571		.037	.017	-.065	.057	-.025	-.043	.471	.342	.084	.343
117.16	16.44	PA	.402	.558		.001	.019	-.003	-.064	.033	.408	.318	.137	.366
118.47	16.87	MA	.519	.378	.511		-.032	-.007	.008	-.039	.304	.459	.087	.343
103.12	16.12	ACS	.305	.336	.372	.299		-.097	.075	.055	.203	.145	.035	.164
115.46	14.86	SM	.503	.306	.497	.656	.341		-.072	.029	.233	.392	.256	.408
116.29	16.98	AI	.255	.187	.358	.483	.033	.418		-.025	.251	.402	.164	.405
120.88	12.75	EI	.401	.259	.347	.562	.090	.648	.459		.355	.287	.199	.396

*The intercorrelations among the ACB tests are presented in the lower half of the 8 × 8 matrix; the residuals after two factors, in the upper half.

TABLE II

CENTROID FACTOR LOADINGS

ACS Tests	Factor I	Factor II	h^2
RV	.684	.190	.504
AR	.614	.462	.590
PA	.698	.199	.527
MA	.788	-.193	.653
ACS	.416	.317	.274
SM	.780	-.251	.671
AI	.519	-.340	.385
EI	.662	-.411	.607
Sums of Squares	3.442	.774	

TABLE III

ROTATED FACTOR LOADINGS

ACB Tests	Factor I	Factor II	h^2
RV	.386	.596	.504
AR	.154	.753	.591
PA	.391	.612	.527
MA	.718	.378	.658
ACS	.102	.513	.274
SM	.751	.329	.672
AI	.614	.089	.385
EI	.768	.131	.607
Sums of Squares	2.382	1.836	

TABLE IV

PROJECTED FACTOR LOADINGS OF THE FOUR COURSE VARIABLES ON THE ROTATED FACTOR STRUCTURE

Course Variables	I Technical	II Academic	h^2
MM	.251	.498	.311
LA	.431	.314	.284
MPD	.237	.045	.058
TCA	.447	.312	.297

TABLE V

PRINCIPAL COMPONENT LOADINGS WITH COMPUTED CENTROID COMMUNALITIES IN THE PRINCIPAL DIAGONAL

ACB Tests	Factor I	Factor II	h^2
RV	.705	.212	.542
AR	.610	.472	.595
PA	.712	.232	.561
MA	.789	-.154	.646
ACS	.407	.347	.286
SM	.801	-.220	.690
AI	.517	-.332	.378
EI	.673	-.408	.619
Sums of Squares	3.526	.791	

titudes among missile technicians suggest another interesting inference. Proficiency differences in various areas of missile technology may be due to extreme emphasis on one or the other aptitudes. The ability to read schematic diagrams, for instance, is probably more closely related to academic aptitude while replacing a malfunctioning component in a given missile system is probably more closely related to technical aptitude. Differential performances in these two activities might well be attributable to differences in aptitude characteristics demonstrable with the ACB variables. A more thorough study of aptitude characteristics with regard to proficiency would probably repay further research in any case; most certainly, such research would clarify the nature of post-graduation job demands that will be required of missile technician trainees.

The sample of technician trainees seems well above average on all eight aptitude measures used in this study. The means and standard deviations obtained on a larger, random sample of enlisted men were presented in a previous study (10:4):

Test	Mean	Standard Deviation
RV	99.4	23.3
AR	94.7	22.5
PA	99.6	24.1
MA	102.5	19.5
ACS	88.6	20.3
SM	100.7	20.7
AI	99.8	20.1
EI	100.5	18.5

An examination of similar data to the left of Table I reveals that the missile technician trainees not only score higher, on the average, than the general population of enlisted men, but they also tend to vary less about the mean than the general population of enlisted men. All in all, the missile technician trainees seem to be above average in every respect with regard to aptitudes.

There is in the results of the present study an interesting anomaly as compared to previous classifications of ACB tests. A previous study (11:19) refers to the Mechanical Aptitude test as a measure of "general aptitude," while the Army Clerical Speed test is referred to as "... a more specific measure peculiar to the job family... ;" otherwise, our Academic factor contains the general measures and the Technical factor contains the more specific measures. True, MA shows some appreciable relation to the Academic factor but it plays a relatively small part in determining the location of this vector; moreover, the contribution of ACS to the Technical factor is negligible. Apparently, the academic and technical aptitude characteristics of missile technicians are sufficiently disparate to locate each of these two tests with others of their more general nature. The MA variable is more closely associated with technical aptitude while the ACS variable has more relationship to general, academic aptitude. As was noted previously, however, the missile technician trainees are above average on both variables as well as all the rest of the ACV variables used in this study.

Aptitude and the Course Variables—Two of the course variables, LA and TCA, each show about twice as much relationship[2] to the Technical factor as compared to the Academic factor. This result appears to be most reasonable. Surely, a technical training program can be expected to exploit the technical talents of its trainees within some amount of academic framework, but it would seem that technical aptitudes should be the most important factor in determining a student's success or failure. From this standpoint, it is encouraging that the total course average, which indicates the student's overall standing, and, in this instance, is made up entirely of written test scores, is determined for the most part by techni-aptitudes rather than academic aptitudes, though the latter receives ample consideration.

Missile Mechanics is the only course variable more closely associated with academic-type aptitudes than with those aptitudes associated with technical prowess. An explanation of this result might reside in the fact that MM is the first portion of the course when students are becoming acquainted with new material; in this circumstance, academic aptitudes might have a larger play in test scores than technical aptitudes. If this is the case, similar results should be obtainable in research with ACB scores and other technical courses. Further studies, of course, might show other reasons for the behavior of MM in this study, but it is not immediately obvious why MM demonstrates about four times as much relationship to academic, general aptitudes than to technical aptitudes. Further research here should prove quite fruitful.

From an inspection of Table IV it seems noteworthy that with the exception of MPD, the factor structure generated by the eight ACB aptitude variables accounts for about 30 percent of each of the course variable's variance. The lack of relationship drawn between the aptitudes and MPD most certainly bears closer scrutiny. Missile Preparation and Depreparation consumes more than half of the instructional time, and receives as much weight in the TCA as both MM and LA together. Still, notwithstanding the importance of the MPD, this study has told us relatively little about the underlying aptitude characteristics associated with this variable. Almost all of MPD's communality is located on the Technical factor, but a more thorough study of this variable should be undertaken in future training research to better understand its nature.

Though the ACB tests proved quite useful in the present study, it should be noted that none of these variables correlated with school grades as well as did the Electronic Placement Test in a previous HumRRO study (1). This result suggests that locally developed tests might well be more closely associated with success in various technical training courses. Certainly, more research should be conducted to combine army-wide aptitude tests with locally built tests in an effort to obtain a better understanding of aptitude characteristics as related to course achievement and job proficiency.

Methodology—Some remarks regarding the methodology in the present study and that in future training research seem to be in order.

Questions regarding the relationships of two sets of variates often arise in training research. We can, of course, canonically correlate the two sets of data and develop the most predictable criterion. Very often, however, as in the present study, we are interested in a parsimonious description of one set of variables, and each of a second set of variable's relation to the first set. The method of factoring the first set of variables and projecting the second set, only makes the resulting structure seem to be a convenient approach in explaining such relationships in the data.

In problems involving differential prediction, we probably cannot do better than distance functions and canonical variates, or perhaps the use of some derived measure such as centour scores (8) related to Mahalanobis' D^2 statistic. But for exploratory studies, such as the present one, the optimum properties of a principal component-type solution do not necessarily provide a clear picture. Kendall (7:28) purports that methods such as the centroid technique "... are objectionable and should not be used when they can be avoided", but we know from Danford's work, for instance, that the centroid, simple summation technique, depending on reflections, bears more relation to analysis of variance than the principal component, weighted summation method. Moreover, it seems most advantageous in exploratory studies to separate variables as much as possible in the factor space, and it appears convenient to accomplish this result via rotated factor solutions.

Certain results, of course, can be supported by principal component solutions. A principal component analysis of the 8 × 8 intercorrelation matrix in Table I, using communalities computed from the centroid solution, produced the results presented in Table V. Again, as in the centroid solution, the first two components together account for approximately 90 percent of the variance. Here, the first component reflects high loadings for all eight ACB tests, while the second component reveals a separation of variables similar to that in the rotated centroid solution. Only about 15 percent of the variance in the principal component solution, however, allows for this differentiation among the variables. The rotated centroid solution, then, appears to provide a greater differentiation among the variables in terms of their variances.

One other problem is that the 8 × 8 intercorrelation matrix was built by averaging 4 matrices of similar correlations. The present writers know of no proof that such a matrix will be Gramian. The requirement of a Gramian matrix, in any case, will carry less import in the centroid technique than in a principal component-type analysis.

Summary

Course grades in a short missile technician's course were studied in relation to aptitude characteristics as reflected in eight Army Classification Battery tests. There were four classes, including a total of 119 enlisted trainees, for whom complete data were available.

The Army Classification Battery tests were factored by the centroid method. Two centroid factors were extracted and rotated graphically. The first factor was determined to be a Technical factor composed of the following tests: Electronic Information, Shop Mechanics, Mechanical Aptitude, and Automotive Information. The second factor was defined as an Academic factor characterized by Arithmetic Reasoning, Pattern Analysis, Reading and Vocabulary, and Army Clerical Speed.

Four written test score distributions were involved in the missile course: one test after each of three parts of the course, and a combination of all three tests to form a total course average. Each of the four distributions were projected separately onto the factorial structure determined by the aptitude variables. The first test covered Missile Mechanic training and was more closely associated with the Academic factor than with the Technical factor. The second test in the course, covering Launcher Area training, and the total course average proved to have twice as much identification with the Technical factor as with the Academic factor. The third test, Missile Preparation, had very little relation to the system of aptitude tests; the attending communality was located, however, on the Technical factor. Suggestions were given for further training research with the Army Classification Battery tests.

The method of the present study was discussed in relation to future training research. The centroid method of factor analysis, and the projection of variables onto the resulting factorial structure, was suggested for further exploratory studies in the area of training.

FOOTNOTES

* Permission is granted for reproduction, translation, publication, use and disposal in whole and in part by or for the United States Government. The opinions expressed are solely those

of the authors and are in no way official; nor are they to be construed as representing those of HumRRO or any other bureau or agency.

•

** The authors wish to express their appreciation to Mrs. Lynne Von Kanel for her most valuable assistance in the computations involved in the present study.
Present address of Dr. Anderson: American Institute for Research, 11607 Washington Place Los Angeles 6, California.

1. This test was built by personnel in the U. S. Army Air Defense School, Fort Bliss, Texas.

2. "Relationship" herein refers to the square of the obtained factor loadings, which indicate directly the proportion of variance involved.

REFERENCES

1. Anderson, H. H., Jr., and J. Whipple, Course Achievement of Students with Unsatisfactory Academic Averages in Basic Electronics, unpublished (Fort Bliss, Texas: U. S. Army Air Defense Human Research Unit.
2. Birnbaum, A. H., and others. Selection and Standardization of Tests for Improved Combat Aptitude Areas, PRB Technical Research Note 87 (December 1957), 14 pp.
3. Campbell, J. T., and others. Procedural Problems in Validating the Army Classification Battery, PRB Research Report 996 (December 1952), 19 pp.
4. Danford, M. B. Factor Analysis and Related Statistical Techniques, Unpublished Ph. D. Thesis, North Carolina State University, 1953.
5. Fruchter, B. Introduction to Factor Analysis (New York: D. Van Nostrand Co., 1954), 280 pp.
6. Guilford, J. P. Fundamental Statistics in Psychology and Education, 3rd Ed., (New York: McGraw-Hill Book Co., 1956), 565 pp.
7. Kendall, M. G. A Course in Multivariate Analysis (New York: Hafner Publishing Co., 1957), 185 pp.
8. Tiedeman, D. V. "Problems and Procedures in Profile Analysis," Proc. Invit. Conf. Tstg Prob., ETS (1953), pp. 54-75.
9. Whitmore, P. G., and W. L. Williams. The Development and Use of a Performance Test as a Basis for Comparing Technicians With and Without Field Experience: The Nike Ajex IFC Maintenance Technician, (Fort Bliss, Texas: U. S. Army Defense Human Research Unit, July 1958, processed.
10. Woods, I. A., and L. K. Burke, Analysis of Army Classification Battery Scores, PRB Technical Research Note 60, December 1956, 11. pp.
11. Zeidner, J., and others. Reconstruction of the Aptitude Areas, PRB Technical Research Report 1095, November 1956, 31 pp.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 3, March 1961)

ESTABLISHING CRITERION GROUPS FOR EVALUATING MEASURES OF CURIOSITY

WALLACE H. MAW
University of Delaware
ETHEL W. MAW
Bryn Mawr College

CURIOSITY HAS long been considered an important attribute, if not always one to be encouraged (4). In recent years it has received increasing attention because of its relationship to creativity. Most schools of psychology implicitly or explicitly indicate an awareness of it (2, 7). Curiosity or exploratory behavior of lower animals has been rigorously investigated (3, 6), and some effort has also been made to study curiosity of human beings (1).

Although a considerable body of literature has accumulated in which curiosity has been discussed, very few empirical studies have been directly concerned with children's curiosity. One reason for the dearth of investigations involving children may be the lack of instruments with which to measure curiosity.

A major problem in developing measures of curiosity is establishing validity. One approach is to identify a group of children of high curiosity and another group of low curiosity and to see whether proposed measures of curiosity discriminate between the two groups. Isolating such criterion groups was the purpose of the present investigation.

Securing Judgments of Children's Curiosity

It was first necessary to define curiosity in terms of behavior. To do this, many statements made by philosophers and psychologists throughout history as well as reports of empirical investigations were analyzed. The analysis is reported elsewhere. **

As a result, for the purposes of this study, an elementary school child was said to exhibit curiosity to the extent that he:

1. reacts positively to new, strange, incongruous, or mysterious elements in his environment by moving toward them, by exploring them or by manipulating them,
2. exhibits a need or a desire to know more about himself and/or his environment,
3. scans his surroundings seeking new experiences,
4. persists in examining and exploring stimuli in order to know more about them.

The teachers of five fifth-grade classes were asked to judge the curiosity of their pupils in light of this definition. At the same time, the pupils were asked to judge their peers and themselves. Thus three judgments of the curiosity of the children of each class were secured—teacher-judgment, peer-judgment, and self-judgment.

Efforts were made to minimize errors inherent in this approach. The evaluative instruments were administered to all classes by one person who used the same procedure in each case. Each class was observed by an independent observer one hour per week for ten weeks. The cumulative record of each child was reviewed and serious discrepancies between the record and the judgments of the evaluators were studied on an individual basis.

Three instruments were developed for the different types of judgments. Each instrument was constructed to include all aspects of the operational definition. In addition, the instruments used by children were either carefully controlled in terms of vocabulary or were read to them.

Each teacher was given the definition of curiosity together with an example for each part of the definition. She was told that all of the kinds of behavior included might not be observable in any one child, but that it was reasonable to suppose that the more of them a child showed, the more curious he was. The teacher was also cautioned that the child who showed the most curiosity might not be the one who was making the best classroom adjustment.

Each teacher was asked to rate her pupils in the following manner:

1. Write the name of the child you consider to have the most curiosity on the first line.

---

*Footnotes will be found at end of article.

2. Write the name of the child you consider to have the least curiosity on the line corresponding to the number of pupils in your class.
3. Next, write the name of the child you would rank second in curiosity on line two.
4. Then write the name of the child you consider to have next to the least curiosity on the line above the name of the child having the least curiosity.
5. Continue ranking in this manner until you have ranked all of the children in your class.

Since this method appeared to be too complicated for the majority of the children to use to evaluate their peers, a Who-Should-Play-the-Part test was created. The children were given the description of eight roles—four of children whose behavior exhibited much curiosity and four whose behavior exhibited little or none. They were told to select children in their classrooms who were generally most like the characters required for the play.

Typical role descriptions are the following:

Part 1. This part will be played by a classmate who keeps working for a long time trying to understand anything new which can be examined. This pupil sticks to problems trying to solve them. This member of the class is the last to give up when the class is looking for answers to questions. This pupil keeps asking questions after everyone else has stopped and will remain working on strange things after others are done. This child often takes things apart, but will work a long time to put them together to find out how they work. ____________

Part 8. This part will be played by a pupil who misses seeing the things that other members of the class see easily. This pupil is not easily disturbed by things that happen in the classroom. When something new or strange is brought into the classroom, this classmate often does not look at it or just gives it a slight glance. ____________

The pupils were asked to evaluate themselves using an instrument consisting of 41 statements regarding habits and attitudes and entitled "About Myself." They were told that there were no "right" or "wrong" answers, that the best answer was what they thought was true of themselves. Although the majority of the items were stated in terms of behavior that logically indicated curiosity, some statements were used that implied just the opposite. Each of the items was related to one or more points of the operational definition.

The following are typical items from the About Myself test:

When there is something new in the room, I notice it right away.

______Never ______Sometimes
______Often ______Always

I like to discover new things.

______Never ______Sometimes
______Often ______Always

I keep away from strange and unusual things.

______Never ______Sometimes
______Often ______Always

When I see a strange machine, I go up to it and look at it.

______Never ______Sometimes
______Often ______Always

All of the judgments of curiosity were made during the first week of March after the teachers and classes had been working together for at least six months. In one classroom the teacher and pupils had been together for sixteen months.

Selecting Groups of High- and Low-Curiosity Children

It was anticipated that certain factors—age, race, popularity, sex, and intelligence might have some effect on the judgment of the evaluators.

Age was kept within narrow limits by restricting the study to one grade in school. The fifth grade was selected to lessen reading problems and to secure children whose interests were less canalized than those of adolescents.

At the beginning of the study, the pupils' ages ranged from ten years, two months to twelve years, eight months. Their Lorge-Thorndike intelligence quotients ranged from 71 to 139.

Of the 158 children, thirteen were Negroes—two girls and eleven boys. The question was raised regarding the extent to which racial bias might influence judgments made by the teachers or the pupils. In one room there was one Negro. In each of the two other rooms there were six Negroes. In one of these, the teacher and the pupils placed the same three children in the upper half of the class and the other three in the lower half. In the other room, the teacher placed five of the children in the upper half while the peers judged four to be in the lower half. By using Fisher's Exact Probability Test, it was determined that this was not significant at the .05 level and therefore should be considered a chance arrangement.

Another question to be answered was whether the substantial agreement found between the judgment of teachers and pupils could be accounted for by popularity. In other words, did teachers and peers tend to rate popular pupils high in curiosity and less popular pupils less high? In order to determine the popularity of the children, the Ohio Social Acceptance Scale was administered. The

TABLE I

AGREEMENT BETWEEN TEACHERS AND PEERS IN RANKING PUPILS IN FIVE CLASSROOMS AS TO THE EXTENT TO WHICH THEY EXHIBIT CURIOSITY WITH THE EFFECT OF POPULARITY PARTIALLED OUT

Room	N	Tau	Level of Significance	Partial Tau
A	31	.37	.0018	.31
B	30	.33	.0052	.30
C	34	.25	.0188	.21
D	30	.42	.0005	.39
E	28	.36	.0035	.35

effect of popularity on the correlation between teacher-judgment and peer-judgment was then partialled out. Table I shows that the effect of popularity was negligible.

It has been rather consistently reported that, in our society, there exists considerable rivalry between the sexes at preadolescence. If sex rivalry existed in these classes, to what extent did it influence pupils' judgment of the curiosity of peers of the opposite sex? To answer this question, the rankings which pupils of each sex gave to like-sex and opposite-sex peers were correlated with each other and with the rankings made by the teachers. The majority of the correlations were moderate to high. All of them were positive.

While these findings indicated that there was considerable agreement among boys, girls, and teachers in judging curiosity, they did not answer the question of whether, on the whole, members of one sex were judged to be more curious than the members of the other. Two approaches were made to answer it. The differences between the means of the scores the peers had assigned to the boys and girls and the differences between the number of boys and girls the teachers had placed in the upper and lower halves of the classes were examined.

Table II shows the mean and variance of peer judgment of curiosity of boys and girls for each class. It shows that in three of the classes the boys were assigned scores that were significantly more variable than those assigned to the girls. In other words, in these rooms there were boys who were scored much higher than the girls in the room. There were also boys who were scored much lower.

Since the differences between some of the variances were significant, the differences between the means were tested with the Mann-Whitney U test. An examination of the means reveals that in some classrooms the girls' means were higher than the boys' means. In others, the opposite was true. In no case was the difference significant.

Although, on the whole, the teachers placed more boys than girls in the upper half of the class and more girls than boys in the lower half, this arrangement was not consistent in all classes as is shown in Table III. When tested, using the chi square technique, the differences were found not to be significant. In light of the evidence it seems safe to assume that the ratings were not unduly influenced by sex preference.

Intelligence was found to be substantially related to both teacher-judgment and peer-judgment of curiosity. Teacher-judgment tended to be more consistently related to intelligence than did pupil-judgment. In all cases, the relationship was positive and significant as is shown in Table IV.

To obtain groups of pupils of high and low curiosity independent of intelligence, lines of regression of curiosity on intelligence were drawn for each class and pupils whose scores deviated from the regression lines were selected. Pupils whose scores were at least one-half standard error of estimate above the regression line were included in the high curiosity group; those whose scores were at least one-half standard error of estimate below were included in the low curiosity group. This was done separately for teacher- and peer-judgment. The high-curiosity and low-curiosity groups established from the two different judg-

TABLE II

NUMBER, MEAN, VARIANCE, AND SIGNIFICANCE OF DIFFERENCE BETWEEN MEANS AND BETWEEN VARIANCES OF PEER-JUDGMENT SCORES OF BOYS AND GIRLS OF FIVE FIFTH-GRADE CLASSES

	Girls			Boys			Means		Variances	
Room	N	Mean	Variance	N	Mean	Variance	U*	P	F	P
A	16	40. 19	69. 40	15	40. 93	357. 13	103. 5	N. S.	5. 17	< . 005
B	19	39. 21	155. 11	11	45. 54	838. 06	91. 0	N. S.	5. 63	< . 005
C	14	46. 07	227. 35	20	35. 50	115. 95	84. 5	N. S.	2. 01	> . 10
D	12	33. 58	164. 08	18	45. 83	592. 92	140. 0	N. S.	3. 51	< . 05
E	10	42. 60	199. 84	18	38. 94	138. 50	78. 5	N. S.	1. 39	> . 10

*The Mann-Whitney U test was employed.

TABLE III

SEX COMPOSITION OF UPPER AND LOWER HALVES OF CLASS ACCORDING TO TEACHER-JUDGMENT OF CURIOSITY, CHI SQUARE, AND LEVEL OF SIGNIFICANCE IN FIVE FIFTH-GRADE CLASSES

Room	Upper Half		Lower Half		Chi Square	Level of Significance
	Boys	Girls	Boys	Girls		
A	7*	9	8	8	.13	.75 > P > .50
B	7	8	4	11	.57	.50 > P > .25
C	12	5	9	9	.81	.50 > P > .25
D	11	5	7	9	1.14	.25 > P > .10
E	11	3	7	8	1.92	.25 > P > .10

*Where the number in the class is odd, the person in the middle is counted with the lower group.

TABLE IV

PRODUCT-MOMENT CORRELATION COEFFICIENTS BETWEEN CURIOSITY AND INTELLIGENCE* OF PUPILS IN FIVE FIFTH-GRADE CLASSES

Room	Teacher-Judgment of Curiosity and I. Q.		Peer-Judgment of Curiosity and I. Q.	
	N	r	N	r
A	31	.67	30	.33
B	28	.60	28	.65
C	30	.51	30	.32
D	30	.43	30	.32
E	29	.57	28	.63

*Intelligence quotients based on Lorge-Thorndike Intelligence Test.

TABLE V

COMPARISON OF HIGH- AND LOW-CURIOSITY GROUPS AS TO AGE, POPULARITY, AND INTELLIGENCE

Item	N	Age		Popularity*		Intelligence	
		Mean	Variance	Mean	Variance	Mean	Variance
Higher	15	130. 33	47. 69	93. 5	523. 18	105. 07	281. 00
Lower	20	128. 30	38. 61	104. 9	449. 09	105. 60	279. 14
		t = . 91	F = 1. 26	t = 1.47	F = 1. 19	t = . 09	F = 1.03
Significance of Difference		P>. 30	P>. 20	P>. 10	P>. 20	P>. 90	P>. 20

*Lower Score = Higher Social Acceptance

TABLE VI

MEAN AND VARIANCE OF HIGH- AND LOW-CURIOSITY GROUPS ON A SELF-APPRAISAL OF CURIOSITY AND SIGNIFICANCE OF DIFFERENCES BETWEEN MEANS AND BETWEEN VARIANCES

Item	N	Mean	Variance
Higher Curiosity	15	73. 53	93. 98
Lower Curiosity	20	62. 55	131. 92
		t = 2. 87	F = 1. 28
Level of Significance		P <. 005*	P >. 20

*One-tail test

ments were then compared. Only pupils who were placed similarly by the two judgments were retained in the tentative criterion groups.

The five regression equations based upon teacher-judgment of curiosity were not significantly different from each other; those based upon peer-judgment were significantly different. In the latter case, a regression equation was written for the five classes as one group. The tentative criterion groups yielded by the use of this equation did not differ from those established by the use of the five separate equations.

There were fifteen children who were judged by both teachers and peers to exhibit more curiosity than would be predicted on the basis of their intelligence quotients. There were also twenty children who were judged by both teachers and peers to exhibit less curiosity than would be expected on the basis of their intelligence quotients. These groups respectively made up the tentative high- and low-curiosity groups. Table V shows that the two groups did not differ significantly in age, popularity, or intelligence.

Questions pertaining to race and sex bias were raised earlier. It seemed pertinent to raise the same questions in regard to the tentative criterion groups. The limited number of Negroes made it impossible to reach definite conclusions, but two Negroes were included in the criterion groups, one in the high curiosity group and one in the low. Ten boys and five girls were in the high curiosity group, and seven boys and thirteen girls were in the low group. When tested, this was found to be a chance arrangement. It seemed likely that race and sex membership had not unduly influenced judgment of curiosity.

Finally, the self-judgments of the children in the tentative criterion groups were examined. On the whole, the children in the high-curiosity group gave themselves higher scores on curiosity than did the children in the low-curiosity group. As shown in Table VI, the difference between the means was significant at the .005 level.

It was concluded that the tentative criterion groups were sufficiently established to be considered the actual criterion groups. They were, therefore, used to validate items intended to measure the curiosity of elementary school children.

Summary

The purpose of the investigation was to establish criterion groups of elementary school children for evaluating measures of curiosity. Teacher-judgment, peer-judgment, and self-judgment of curiosity of children in five fifth-grade classes were obtained after the classes had been working together for six months.

The judgments of teachers and peers were not significantly affected by race, sex, or popularity, but were significantly related to intelligence. With statistical control of intelligence, high- and low-curiosity groups were selected on the basis of teacher- and peer-judgment. On a self-appraisal of curiosity the children of the high-curiosity group rated themselves significantly higher in curiosity than did the children of the low-curiosity group. The final criterion groups consisted of fifteen children of high curiosity and twenty of low curiosity.

FOOTNOTES

* The research reported herein was performed pursuant to a contract with the United States Office of Education, Department of Health, Education and Welfare.

**A 29-page report describing this analysis and giving complete data in tables has been deposited with the American Documentation Institute. Order Document No. 6558, remitting $2.00 for 35 mm. microfilm or $3.75 for 6 × 8 in. photocopies. Advance payment is required.

REFERENCES

1. Berlyne, D. E. "An Experimental Study of Human Curiosity," British Journal of Psychology, XLV (1954), pp. 365-73.
2. Brill, A. A. (Ed.) The Basic Writings of Sigmund Freud (New York: Modern Library, 1938).
3. Carr, R. M., and Williams, C. D. "Exploratory Behavior of Three Strains of Rats," Journal of Comparative and Physiological Psychology, L (1957), pp. 621-23.
4. Ecclesiasticus III:23.
5. Emch, M. "On 'the Need to Know' as Related to Identification and Acting Out," International Journal of Psycho-Analysis, XXV (1944), pp. 13-19.
6. Harlow, H. F. "Learning a Satiation of Response in Intrinsically Motivated Complex Puzzle Performance by Monkeys," Journal of Comparative and Physiological Psychology, XLIII (1950), pp. 289-94.
7. McDougall, W. An Introduction to Social Psychology (London: Methuen, 1908).

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 3, March 1961)

AN INVESTIGATION OF THE TEACHING OF CHRONOLOGY IN THE SIXTH GRADE

VAL E. ARNSDORF
University of California, Berkeley

THE TEXTBOOKS basic to the teaching and learning of the social studies are replete with words and phrases relating to time and historical chronology. Despite the frequent occurrence of such terms, little is known about children's ability to comprehend them and to use them effectively in dealing with the problem of this subject matter field. Moreover, little is known--perhaps even less--about instructional materials and methods which may increase understanding of these terms.

This article reports one attempt to secure research information in this relatively neglected area. The investigation involved the use of two comparable groups in Grade 6. While both groups studied the same social studies unit for seven weeks, namely, "Ancient Civilizations," no effort was made to assist the teachers and the pupils in the control group. The teachers and the pupils of the experimental group, on the other hand, had the opportunity to try out the effect of specific aids to be described shortly. It is to be borne in mind, however, that prior to their use in this experiment, there was little empirical evidence that these aids would be really beneficial.

Previous Related Research

There is in general no lack of research on the factors that influence children's ability to understand subject matter and to cope with the problems therein; but the research which relates specifically to understanding and problem solving in subject matter making large use of time and historical chronology is very limited indeed. The three most pertinent research reports are reviewed very briefly.

Oakden and Sturt (1922) (5) investigated the developmental process in acquiring understanding of chronology concepts, using a series of tests. They concluded from their test results and from their observations that growth in understanding of time is slow and that an adult level of understanding is not attained before the age of thirteen or fourteen. While they were able to demonstrate differences in rates of development, they did not set up any particular program of instruction designed to accelerate development.

Pistor (1940) (6) studied the effects of training on the development of time concepts. His method was to compare the results of two equivalent groups, one following a separate subject approach in history and geography in grades four and five, the other following, basically, a geographical approach with history taught incidentally. The mean scores of the two groups on the final tests of time concepts were about equal. Greater growth, but not significantly so, was reported after repeating the experiment with a similar program in grade six. According to Pistor, his evidence indicated that "maturation" is dominant over training in the development of time concepts. There seems to be, in his data, little support for this hypothesis.

A third study, Friedman's (1944)(3), dealt with the variety and importance of time concepts in the life of both children and adults. Using interviews, tests, and time-lines, Friedman found that children have a slight understanding of time when they enter school and progress in understanding with each succeeding grade, attaining full comprehension of our conventional time system by the sixth grade and approaching maturity in comprehension of time concepts by the tenth grade. The relationships reported between scores on tests dealing with time on the one hand and intelligence on the other were low, as were also the relationships between test scores and socio-economic status. Differences in performance on the tests by boys

---

[*] Abstract of an unpublished dissertation by the same title, offered as partial fulfillment of the requirements for the Ph.D. degree at the University of Minnesota, 1959.

and girls were not significant.

As a group, these studies agree in emphasizing the difficulties involved in understanding time concepts; also in concluding that maturity in such understanding is the result of a slow developmental process. However, none of the three provides much evidence on the effect of organized instructional attempts to increase children's ability to understand time concepts as they are employed in the social studies. It was for this reason that the present investigation was undertaken.

The Present Study

Subjects. The experiment proper was delayed until the initial testing program could be completed. At this time, the Lorge-Thorndike Intelligence Test, Non-Verbal Battery, Level III, was administered. (The scores were later used also to divide the groups into three ability levels, from top to bottom, 110-145, 95-109, and 50-94.) Table I summarizes data on sex, average class size, and average I. Q. of the experimental and control groups, showing them to have been closely comparable on the factors mentioned. All subjects were drawn from the same mid-western city school system.

Differences in Instruction. As has been stated above, both experimental and control groups studied the same social studies unit from the same textbooks and for the same period of time. The instructional program of the experimental group, unlike that of the control group, was intended to identify ahead of time certain anticipated learning difficulties and to provide assistance in meeting them. It is to be re-emphasized, however, that the devices and suggestions offered had not been previously tested in order to determine their worth.

In general terms, the forms of special assistance given the experimental classes in the seven-week period were as follows:

1. Specific identification of, and instruction in, every term in the texts relating to time.
2. Use of various time-lines and charts to add concreteness to such abstract terms.
3. The writing of biographical and autobiographical sketches to stimulate interest in past and present events.

Testing Instruments Employed. Reference has already been made to the uses of 1) the Lorge-Thorndike Intelligence Test. 2) The comprehension section of the Gates Survey Reading Test was used to control reading ability as a factor in later analyses. 3) Gains in the abilities essential to success in the social studies were determined by administering, both initially and finally, the Iowa Every-Pupil Test B, Basic Work-Study Skills.

4. In order to measure increases in the understanding of historical time, the writer prepared an original test battery which was given both groups of pupils before and after the experimental period. Included were the following six sub-tests: Vocabulary of Chronology, Ordering Four Events without Dates, Ordering Two Events without Dates, Relative Time, Ordering Four Events with Dates, and Time Absurdities. The items included in this battery were selected from information found in reference books, basal texts, and supplemental social studies books available to pupils in the intermediate grades. Reliability coefficients for the six tests, based on the results of the initial testing program, ranged from .98 to .61 and indicate some variation in the consistency of these measures. Inter-correlations among the tests ranged from .581 to .014 and reveal that the tests did differentiate fairly well various ways of measuring the understanding of historical time.

Hypotheses Tested. The effects of the differentiated programs of instruction were measured principally with respect a) to gains in basic study skills and b) to gains in understanding of time, as determined by scores on the six separate tests. In each type of comparison, there was interest in the sex of the learners and in their levels of intellectual ability. The method of analysis adopted may be illustrated with respect to gains in basic study skills, which led to a total of four hypotheses, since sometimes additional controls (e.g., reading comprehension) were employed, and sometimes they were not. The four hypotheses for basic study skills are:

1. There is no significant difference in the final mean scores on the tests of basic study skills a) for the experimental and for the control groups; b) for boys and for girls; and c) for levels of intellectual ability; nor is there d) a significant interaction between treatment* and sex or level of intellectual ability.
2. There is no significant difference in the final mean scores on the tests of basic study skills, when the initial scores on these tests are controlled a) for the experimental and for the control groups; b) for boys and for girls; and c) for levels of intellectual ability, nor is there d) a significant interaction between treatment* and sex or level of intellectual ability.
3. There is no significant difference in the final mean scores on the tests of basic study skills when reading comprehension scores are controlled a) for the experimental and for the control groups; b) for boys and for girls; and c) for levels of intellectual ability, nor is there d) a significant interaction between treatment* and sex or level of intellectual ability.
4. There is no significant difference in the

---

* The single word "treatment" is used here, in the interest of brevity, to refer to method of instruction, differentiated as previously explained for the experimental and the control groups.

TABLE I

SEX, CLASS SIZE, AND MEAN I. Q. OF SAMPLE

Group	N	Boys	Girls	Class Size	Mean I. Q.
Experimental	322	166	156	32. 2	101.01
Control	241	121	120	30. 0	99. 23
Total	563	287	276	31. 2	100. 82

TABLE II

BASIC STUDY SKILLS: SUMMARY OF F-VALUES OF THE ANALYSIS OF VARIANCE AND CO-VARIANCE TESTS OF THE SIGNIFICANCE* OF DIFFERENCES BETWEEN THE MEANS OF THE EXPERIMENTAL AND CONTROL GROUPS

Group		F
Experimental and Control	No other controls	1. 27
Experimental and Control	Initial scores (basic skills) controlled	1. 19
Experimental and Control	Reading comprehension scores controlled	1. 60
Experimental and Control	Initial scores (basic skills and reading comprehension both controlled)	2. 89

* None of the differences is significant.

final mean scores on the tests of basic study skills when both the initial scores on these tests and on reading comprehension are controlled a) for the experimental and for the control groups; b) for boys and for girls; and c) for levels of intellectual ability; nor is there d) a significant interaction between treatment and sex or level of intellectual ability.

No single score was computed for understanding of time and historical chronology. Instead, the gains on each of the six sub-tests are treated separately. For each of these, then, there are four null hypotheses corresponding to the four for basic study skills. The first null hypothesis for study skills needs merely to be re-worded, to get six hypotheses relating to understanding of time, by substituting in turn Vocabulary, 4-Event Order, etc., for the term study skills. There is therefore no reason to state in full the resulting twenty-four null hypotheses, expecially since their nature will be clear enough when the findings are presented.

Statistical Procedures. All the null hypotheses were tested by an approximate method of analysis of variance and covariance. The unequal frequencies resulting from the multiple classification of the data prevented the use of a more exact test. The robust character of the basic technique of analysis of covariance (1, 2, 4) provides adequate defense for the techniques used, even though not all assumptions were satisfied. This fact, however, makes it necessary to be somewhat guarded in interpreting results.

Gross Findings--Basic Study Skills. At this point, only the data relevant to the comparisons labeled a) in the four null hypotheses are treated, those for comparisons b), c), and d) being postponed for the time being. In other words, the findings here in question relate to the experimental and the control groups as wholes, under the four conditions described. The quantitative facts are assembled in Table II. None of the differences between the two groups so far as basic study skills are concerned proved to be significant, whether the comparisons are based upon the test scores alone or on the test scores with three type types of control: 1) initial scores on the tests, 2) reading comprehension scores, and 3) both 1) and 2). What this means is that the instruction designed to increase the ability to deal with time concepts had no discernible effects upon study skills as such. Perhaps none should have been anticipated since the types of training given the experimental group were specialized and could hardly be expected to develop skills either that are represented as such in the study skills test or that could be generalized to reveal their presence through transfer.

Sub-abilities in Understanding Historical Time. Table III, like Table II, contains the results of comparing relative gains for the experimental and the control groups as wholes. For each of the six tests of time understanding there are four null hypotheses, as already explained, to make a total of twenty-four. In all twenty-four instances, the advantage lay with the experimental group, but in two, by margins so small as to be completely negligible (2-event order: no dates). In the remaining twenty-two cases, all except four differences are significant.

Of the sub-abilities measured, one produced no significant differences at all in the relative gains made by the two groups. The corresponding test required pupils to arrange in correct temporal sequence two undated historical events. Why the instruction on the understanding of time given to the experimental group produced no effects at this point is not known.

Of the remaining five sub-abilities, the one making the poorest record (that is, favoring the experimental group least often) was 4-event order: no dates; and its similarity in psychological demands to the least successful of all (2-event order: no dates) is obvious. The experimental subjects surpassed the control subjects only a) when there were no controls at all (beyond that introduced by the differentiated instruction itself) and b) when reading comprehension scores were statistically controlled.

Marked success attended efforts to teach children in the experimental group a) the vocabulary of historical time, b) the ability to distinguish between the relative times of events, c) skill in placing four dated events in correct temporal order, and d) the ability to detect time absurdities. All sixteen potential differences are significant. Since care was taken to prevent teaching which would amount to coaching for the tests, it is believed that the superiorities of the experimental children in these four sub-abilities are realities and not artifacts.

Findings Related to Sex and Ability Sub-groups. In this section of the report it will be possible to include only part of the data obtained when experimental and control groups were treated not as wholes, but in terms of sex sub-tests and of sub-groups made upon the basis of intell egence test scores. There were three sub-groups of the latter kind.

In Table IV are presented the F values obtained when, with the initial scores controlled in each case, comparisons were made of the relative gains on the seven criteria--basic study skills, and the six aspects of time understanding taught the experimental group and tested in both groups before and at the conclusion of the experimental period.

According to the figures in the upper row of the table, the topmost of the three intelligence groups, whose scores had surpassed those of the middle and lowest groups on the preliminary tests, maintained their superiority and added to

TABLE III

SIX SUB-ABILITIES, UNDERSTANDING OF TIME: SUMMARY OF F-VALUES OF THE ANALYSIS OF VARIANCE AND CO-VARIANCE TESTS OF THE SIGNIFICANCE OF DIFFERENCES BETWEEN MEANS OF THE EXPERIMENTAL AND CONTROL GROUPS

Group	Vocabulary	4-event order: no dates	2-event order: no dates	Relative time	4-event order: dates	Time absurdities
E and C; no other controls	26.89*	11.47*	2.25	38.16*	40.99*	18.95*
E and C; initial scores controlled	17.33*	2.67	--	43.75*	50.50*	22.91*
E and C; reading comprehension scores controlled	28.55*	11.43*	2.16	39.12*	45.08*	21.97*
E and C; both initial and reading scores controlled	19.15*	3.18	--	44.10*	52.97*	25.56*

* Differences significant at the .01 level.

TABLE IV

SUB-GROUPS, SEX AND INTELLECTUAL LEVEL: SUMMARY OF F-VALUES OF THE ANALYSIS OF COVARIANCE TESTS OF THE SIGNIFICANCE OF DIFFERENCES, INITIAL SCORES IN EACH OF THE SEVEN BASES OF COMPARISON CONTROLLED

Sub-group	Basic study skills	Vocabulary	4-event order: no dates	2-event order: no dates	Relative time	4-event order: dates	Time absurdities
Level of ability	19.69*	9.76*	11.83*	7.62*	3.64	32.98*	20.37*
Sex	5.33	--	--	515.04*	--	--	--

* Differences significant at the .01 level.

it by significant differences in all but one of the seven types of measure, namely, relative time.

According to the figures in the second row, there was enough difference to be worth reporting only in the case of basic study skills and of the ordering of two undated events. In both instances superiority lay with the boys, but in only one of them, significantly so.

An attempt was made to analyze the "interaction" between method of instruction experienced on the one hand and sex and intellectual level on the other. (This type of analysis is represented in item d) in the four null hypotheses stated on a previous page. When, as in Table III, initial scores were controlled in all analyses, only one significant difference was found. This had to do with the ordering of two undated events, and it was accounted for by the excessive influence of the boys. The final mean scores on this criterion were higher a) for boys of superior ability, b) for boys in the experimental group, and c) for boys of superior intellectual ability in the experimental group.

Discussion

Interpretation of the findings, to be summarized below, is dependent upon the conditions under which the investigation was made. a) A prime consideration is the validity and the reliability of the measures employed. Some information has been given above concerning these matters, and it is believed that the quality of the measures is adequate for the purposes in question. b) All subjects were drawn from self-contained classrooms in grade six in a single mid-western community. How far the present findings can be generalized for samples of a different character, it is impossible to say.

c) No controls were placed on the subjects' previous educational achievement and experiences in the social studies or in any other area of the curriculum; that is, no controls beyond those of random selection and statistical analysis. d) The instructional program used with the experimental subjects had few provisions for caring for individual differences. It probably would have proved much more effective with improvements at this point.

e) The vocabulary taught the children was confined to the time-terms of the basal social studies textbook used by both groups of children. f) The experiment ran for only seven weeks -- relatively long perhaps as such experiments go, but short when considered educationally. What effects of the experimental program would be with a longer period for learning is problematic. g) After all, but a single instructional program was tried out, and this without much prior evidence concerning the probable effectiveness of its various aspects. What might be accomplished by a different, or a better program, is also problematic.

The findings of the investigation may be summarized as follows:

1. The instructional program, which for seven weeks emphasized the identification of, and specific instruction on terms relating to chronology and which made use of time lines and biographical and autobiographical materials prepared by the children, fostered a) the comprehension of definite and indefinite time-terms; b) the ability to recognize relative lengths of time between periods and to ascertain the similarity of time-distances with reference to given events; c) skill in ordering events with dates, and d) competence in recognizing time absurdities.
2. This instructional program had little or no effect on other study skills as measured by an appropriate test.
3. Its usefulness in teaching the ability to order undated events varied with the number of events to be placed in temporal sequence. It helped the more capable boys most when only two undated events were involved.

Two additional observations seem to be warranted. The first is that, if the evidence adduced is trustworthy, we can say with confidence that children can profit from systematic instruction of the kind undertaken, to increase considerably their understanding of and ability to use the time relationships common in the social studies. This judgment is in contrast with Pistor's conclusion as stated above, which was to the effect that we can only wait for time to bring the desired changes--the skills and knowledge involved are the product of maturation.

The second observation is based upon much evidence, which can be only hinted at in this report, that children encounter a good deal of difficulty in the social studies because of their deficiencies in dealing with time relationships. The success of the program set up for this study implies the value of other investigations with other plans of instruction, including devices and aids not used in this research. Long-time longitudinal inquiries starting in grade one should also provide a good deal of information for the improvement of instruction at the points of interest in the research described above.

REFERENCES

1. Cochran, W. G. "Some Consequences When the Assumptions Underlying the Analysis of Variance Are Not Satisfied," Biometrics, 1947. 3, pp. 22-38.
2. Eisenhart, Churchill. "The Assumptions Underlying the Analysis of Variance," Biometrics, 1947, 3, pp. 1-21.
3. Friedman, Kopple C. "The Growth of Time Concepts," Social Education, 1944, 8, pp. 29-31.

4. Graybill, Franklin. "Variance Heterogeneity in a Randomized Block Design," Biometrics, 1954, 10, pp. 516-520.

5. Oakden, E. C. and Sturt, Mary. "Development of the Knowledge of Time in Children," British Journal of Psychology, 1922 12, pp. 309-336.

6. Pistor, Frederick. "How Time Concepts Are Acquired by Children," Educational Method 1940, 20, pp. 107-112.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 3, March 1961)

THE RELATION BETWEEN ARITHMETIC RESEARCH AND THE CONTENT OF ARITHMETIC TEXTBOOKS, 1900-1957

MOTHER M. CONSTANCE DOOLEY, R. S. H. M.
Marymount College, Palos Verdes Estates, California

THE PLACE GIVEN to arithmetic in the elementary school instructional program has varied from time to time. At the end of the colonial period arithmetic had just emerged from occupying, as described by Jessup (2), an incidental place with no formal recognition of the subject in the time allotment in the daily program. However, the rapid development of commercial interests soon brought out the practical values of this subject and toward the end of the nineteenth century, half the school day was often devoted to arithmetic.

During the last decade of the nineteenth century, many discussions were held concerning the place that arithmetic should occupy in the elementary curriculum and the proper content that should be taught. The content and method of arithmetic prior to 1900 had been directed toward 1) mastery of the fundamental processes, 2) the extension of the basic facts to the higher decades, 3) the memorization of formulas, and 4) the application of these formulas in as many different situations as time would permit. Scientific studies were begun about 1900 and the years since then have witnessed great changes in the development of methods of teaching arithmetic.

Examination of professional literature and elementary guidebooks on the subject now reveals several fundamental changes in emphasis with respect to the purpose and the place of arithmetic as a school subject. These sources indicate that the aims of arithmetic still point toward mastery of the basic skills but that arithmetic goes beyond this mastery to include 1) a meaningful understanding of numbers and the number processes, 2) the development of the ability to generalize, and 3) the application of arithmetical principles in real life situations.

In 1923, David Eugene Smith evaluated the progress that had been made in these words:

"It was about twenty-five years ago that the movement for weighing values in arithmetic seriously began in this country. It was then that works of promise upon the teaching of the subject began to appear, and it was only a little later that the textbooks first attacked successfully the unreal topics and problems of the past and set a standard for the real ones of the present and the future. Since that movement began there has been a veritable revolution in the subject matter, in its arrangement, in the spirit with which it is presented, and in the textbooks in which the work is set forth." (5)

Research studies continued and newer editions of textbooks were published. In 1950, Wilburn and Wingo stated:

"In spite of the research and publications of findings in the psychology of arithmetic and related fields, and in spite of a large number of improvements in materials for teaching, the program of arithmetic in many elementary schools is not significantly different from the program of schools in the 1890's. There are superficial differences, to be sure, but in fundamental respects the same procedures and the same psychology of learning often obtain." (6)

There is presently available an immense amount of critical, scientific material bearing upon the appropriate content and the best methods of teaching arithmetic.

The textbook was chosen for analysis in this study as a source of the application of research recommendations. Reeder (4) in his historical survey stated that "the best expression of the methods of teaching any branch of the curriculum at any period of its history is revealed in the textbooks of that period." Judd (3) held that "there is no influence in American schools which does more to determine what is taught to pupils than the textbook," while Bagley (1) referred to the textbook as the principal agency of instruction in American schools.

This study was set up to discover the trends that could be found in the acceptance or non-acceptance of research recommendations in the teaching of arithmetic, as indicated by a study of the content of elementary arithmetic textbooks published in the United States between 1900 and 1957. A first purpose was to note the degree of connection between the recommendations made by men who had carried on scientific investigations in arithmetic and the changes that had been introduced into the content of arithmetic textbooks. A second purpose was to check the recommendations which seemed to have been rejected by the authors of textbooks although presented as substantiated, credible findings based upon scientific investigation.

Behind these two purposes were the ultimate intentions of 1) determining the status of arithmetic research in the curriculum development program, and 2) isolating, if possible, reasons for the gap between recommendations for changes in teaching methods and the application of those recommendations.

The following questions were proposed: 1) Are the reasonably credible conclusions of research studies finding their way into textbooks? If so, at what rate? 2) Can reasons be identified which might explain the acceptance of some and the nonacceptance of other credible research findings? 3) Do the new textbooks generally recommend those techniques which have been tried out experimentally and found successful? 4) Is there any evidence that some techniques are recommended by textbook authors before they have been tried out experimentally? 5) Is there a time pattern which is generally followed between the recommendation of research studies and their application in textbooks?

An over-all review of the available literature on the research studies in elementary arithmetic was made to determine the trends that had developed since 1900 in content and in method. Two lists were drawn up, one with topics which seemed to have been definitely influenced by research recommendations, the other containing topics which apparently had not yet been greatly influenced by research. Twenty-five topics were thus gathered together and presented to a national jury of nineteen experts who chose twelve for this study.

The topics which were selected by the jurors are listed here with the frequency of choice.

LIST I--Topics that apparently have been influenced by research recommendations.

1.	Method of placing the decimal point in the quotient.	15
2.	Apparent method vs. increase-by-one method of determining the quotient.	13
3.	Imaginative settings for verbal problems.	10
4.	Placement of long division.	10
5.	Elimination of awkward and unrealistic fractions.	9
6.	Placement of common fractions.	9

LIST II--Topics that apparently have not yet shown the influence of research recommendations.

1.	Testing for concepts rather than for speed and accuracy only.	19
2.	Developing and extending generalizations rather than memorizing rules.	17
3.	Building concepts through the use of concrete materials in classes above the primary grades.	13
4.	The use of illustrations as visual aids rather than as decorations.	11
5.	Method of placing the decimal point in the quotient.	10
6.	Rationalizing division of fractions through using the common denominator rather than the inversion method.	10

A survey was made to locate the research recommendations and the critical opinion on each of these topics. This was followed by the examination of 153 series of elementary arithmetic textbooks published in the United States by twenty-nine different publishers between 1900 and 1957. Trends were indicated in two ways. Tables were compiled showing the dates when the application of research recommendations was noted in each of the different series. Then figures were drawn which showed the percentage of application at different periods of research activity.

Summary and Conclusions

Accepted Findings. The effect of research upon the content presented and the methods suggested in arithmetic textbooks was found to have been direct and immediate in nine of the twelve topics investigated.

1. Between 1900 and 1917, 69.2 percent of the textbooks examined recommended using the integer method for determining the placement of the decimal point in the quotient. From 1917 to 1946, all the textbooks presented this method but since 1946, 93.7 percent recommend it while 6.3 percent suggest teaching the older, subtractive method.

2. The question of the choice between the apparent and the increase-by-one method of estimating the quotient has never been settled but textbooks have followed research recommendations. From 1900 to 1927, 51 percent of the authors recommended the apparent method; between 1928 and 1940, 71.4 percent advocated the increase-by-one method. Since 1940, the trend has been slightly in favor of the apparent method with 57.5 percent recommending this method, for research has also shifted back to it.

3. Imaginative settings have been accepted as the bases for verbal problems. Between 1900 and 1919, 79.1 percent of the books examined

made no use of verbal settings; from 1920 to 1935, 58.6 percent made great use of such settings while only 3.4 percent did not use them at all. Since 1935, 90.0 percent of the books examined were found to make great use while 1.25 percent made no use of imaginative settings for verbal problems.

4. Awkward and unrealistic fractions also showed very definite trends. When the books published between 1900 and 1920 were examined, it was found that only 17.5 percent had eliminated such fractions for the most part, while 42.5 percent made no effort to avoid them. Between 1920 and 1930, 70.0 percent of the textbooks examined had eliminated them and since 1930, 94.4 percent have been free from the large, awkward, and unrealistic fractions that were once widely used.

5. Grade placement of long division and of common fractions has been raised. Examination of books published between 1900 and 1906 showed that 61.5 percent of the authors introduced the process of long division in third grade. Between 1907 and 1930, it was raised to fourth grade in 88.2 percent; from 1931 to 1940, 50.0 percent of the texts introduced it in the fourth, 30.0 in the fifth, and 20.0 percent spread it out over both grades. From 1941 to 1957, 81.5 percent have placed it in the fifth grade.

Between 1900 and 1910, 36.8 percent of the textbooks examined presented addition and subtraction of common fractions in third grade and 47.4 percent introduced these two processes at fourth grade level. From 1910 to 1930, 37.9 percent advocated the fourth grade while the remaining 62.1 percent divided the teaching beween fourth and fifth grades. From 1930 to 1940, 30 percent had moved the placement up to the fifth grade and since 1940, 50.0 percent have favored fifth grade with 44.7 percent dividing the work between fourth and fifth, leaving only 5.3 percent still presenting these two processes completely in the fourth grade.

Multiplication and division of fractions set a more definite trend. From 1900 to 1925, 17.6 percent of the textbooks examined set the fourth grade while 67.6 percent set the fifth grade as the level for introducing these topics. Between 1925 and 1940, the fourth grade was eliminated completely, 47.8 percent set the fifth and 30.4 introduced the work in the sixth grade. Since 1940, 67.6 percent begin instruction in these two processes in the sixth grade.

6. Illustrations are now used as visual aids. Between 1900 and 1937 when the earliest research recommendations appeared, 30.4 percent of the textbooks examined had no illustrations of any kind while only 16.1 percent made use of illustrations as visual aids. Between 1937 and 1946, all the texts examined contained illustrations but 47.1 percent were decorative only, 23.5 percent had diagrams only, and 29.4 percent used the illustrations as visual aids. Since 1947, 80.9 percent of the illustrations found in the textbooks examined were used as visual aids.

7. The common denominator method is used to introduce the process of division of fractions but is not presented as the basic method in any of the recent textbooks examined. From 1900 to 1927, 55.5 percent of the books examined used the inversion method, 24.4 percent the common denominator method, and 20.0 percent the reciprocal method. Between 1927 and 1941, all the books examined presented only the inversion method. Since 1942, 75.7 percent have advocated the inversion method while 24.3 percent introduce the topic by the common denominator method but change to the inversion method almost immediately.

8. The inclusion of tests designed to measure concepts was found to be a recent trend. From 1900 to 1930. 97.0 percent of the texts examined did not include such questions. Between 1930 and 1940, this figure dropped to 52.6 percent as 31.5 percent contained some questions and 15.9 percent made considerable use of questions to measure concepts. Since 1940, 46.0 percent of the books examined have made considerable provision for them while 26.0 percent did not have any of these questions.

9. The use of materials that should contribute to the development of generalizations was found to be a very recent trend. From 1900 to 1945, 92.5 percent of the texts examined made no provision for such materials. Since 1945 when the Second Report of the Commission on Post-War Plans stressed the need for this type of material, 56.3 percent of the texts have suggested considerable use, 21.8 percent some use, and 21.8 percent have not advocated the use of this type of material.

Rejected Findings. Recommendations presented in research studies were found to have been modified or rejected by authors of textbooks in three of the topics under investigation in this study.

1. The major research studies which were carried on from 1926 to 1938 recommended that the introduction of the process of dividing by two or more figures be delayed until the sixth grade and spread out over three years. Modification of this recommendation was noted when it was found that 81.5 percent of the textbooks published since 1930 introduced the topic in the fifth grade and that none introduced it in the sixth or expected the process to take more than two years for the complete presentation.

2. Grade placement of common fractions followed a similar trend. The major research studies recommended that these processes be spread out from the fifth through the ninth grades, and suggested that some phases should never be taught. It was found that addition and subtraction

of common fractions have been placed in the fourth grade by 5.3 percent of the texts examined, in the fourth and fifth grades by 44.7 percent, and in the fifth grade by 50.0 percent of the textbook authors since 1940. Multiplication and division of common fractions have been introduced in the sixth grade by 67.6 percent of the authors but none delay it until the seventh. All of the processes are completely presented by the end of the sixth grade.

3. The use of concrete materials in classes above the primary grades has not been strongly influenced by research findings as yet. From 1900 to 1930, 96.5 percent of the texts examined did not recommend the use of concrete materials above the primary grades. Between 1930 and 1945 many articles advocating such use appeared in professional journals and the examination of textbooks showed that 82.8 percent still did not make any reference to these materials but that 17.1 percent did. Since 1940, 39.5 percent of the texts examined have advocated the use of concrete materials in classes above the primary grades. However, no research on this topic was published until 1950. The few studies that were located have reported that while such use has not retarded learning, no significant differences could be found. Examination of textbooks indicates that the trend to include such materials has continued to increase strongly since 1954.

The Status of Research

The role of research in the arithmetic curriculum development program has been an important one.

1. Many studies were designed to evaluate practices which had been in use for many years.

2. When the recommendations were clear, concise and exact, they were incorporated into some textbooks within five years.

3. When the recommendations were general, intangible, or based upon subjective data, they were not applied as rapidly as where research findings were precise and well supported by adequate data. The lack of clearness and explicitness in the presentation of the recommendations led to slowly developing trends.

4. Those recommendations which were published in the yearbooks of the National Society for the Study of Education and the National Council of Teachers of Mathematics tended to be applied very quickly.

REFERENCES

1. Bagley, William C. "The Textbook in American Education," Educational Administration and Supervision, 17: 241-53, April, 1931.
2. Jessup, W. A., and Coffman, L. D. The Supervision of Arithmetic, New York: The Macmillan Company, 1916, 219 pp.
3. Judd, C. H. "Analyzing Textbooks," Elementary School Journal, 19: 143-54, October, 1918.
4. Reeder, Rudolph R. Historical Development of School Readers and Methods, New York: Columbia University Press, 1900, 93 pp.
5. Smith, David Eugene. The Progress of Arithmetic in the Last Quarter of a Century, Boston: Ginn and Company, 1923, 93 pp.
6. Wilburn, D. Banks, and Wingo, G. Max. "In-Service Development of Teachers of Arithmetic," The Teaching of Arithmetic, pp. 251-68. Fiftieth Yearbook of the National Society for the Study of Education, Part II. Chicago: University of Chicago Press, 1951.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 3, March 1961)

A QUICK - BUT ACCURATE - APPROXIMATION TO THE STANDARD DEVIATION OF A DISTRIBUTION

ROBERT L. LATHROP
University of Minnesota

Introduction

TO THE person familiar with the procedure, and having access to a desk calculator, the computation of the standard deviation of a distribution is not difficult. There are, however, many people who do not have one or both of these requisites but would still find the standard deviation a useful quantity if it could be easily and accurately approximated. Teachers, for example, find the standard score an appealing unit for expressing test scores but will frequently not take the time or trouble to compute the standard deviation for that purpose alone. It was, in fact, this particular problem that led the writer to examine various procedures for approximating the standard deviation of a distribution of test scores.

The Rationale

After examining a number of alternative procedures, a method for approximating the standard deviation mentioned briefly by Diederich (1) seemed to hold considerable promise. This method expresses the variability of a distribution as the difference between two quantities drawn from the tails of the distribution. The logic of this approach is much the same as the argument underlying the use of the range or the quartile deviation as a measure of variability. That is, for two distributions, the one having the greater variability will have the greater difference between two comparable points on the distribution.

In symbolic terms, this approximation of the standard deviation of a finite distribution can be expressed as:

$$\hat{s} = \frac{K}{N}\left[\Sigma X_u - \Sigma X_l\right] \quad . \qquad (1)$$

Where $\hat{s}$ is an approximation of the standard deviation $s = \sqrt{\frac{\Sigma(X - \overline{X})^2}{N}}$

K is a constant determined by the proportion selected for the upper or lower groups (the method for computing K will be described in a following paragraph).

N is the total number of values in the distribution.

ΣX_u is the sum of scores in the upper portion of the distribution.

ΣX_l is the sum of scores in the lower portion of the distribution.

For initial convenience, consider X as a variable which is normally distributed. In this case:

$$\Sigma X_u = Np\left[\, s\left[\frac{y}{p}\right] + \overline{X}\right] \quad . \qquad (2)$$

Where p is an arbitrarily selected portion of the distribution to be included in the upper (or lower) group of scores.

y is the ordinate height at the point where the upper scores are divided from the remainder of the distribution.

$\overline{X}$ and s are the mean and standard deviation of the distribution of X.[1]

Similarly

$$\Sigma X_l = Np\left[\, s\left[\frac{-y}{p}\right] + \overline{X}\right] \quad . \qquad (3)$$

If not apparent to the reader, $\left[\frac{y}{p}\right]$ represents the average standard score in the upper group of scores,[2] Np the number of values (scores) in either extreme group, and s and $\overline{X}$ the quantities needed to convert the standard scores to raw

---

[1] For this finite example, the author has chosen to use the symbols $\overline{X}$ and s rather than the Greek letters μ and σ commonly associated with the infinite normal distribution.

[2] For a discussion of mean sigma distances, see Wert (2:67).

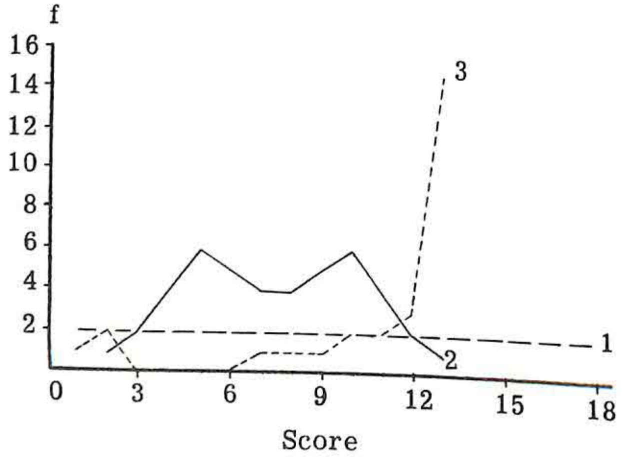
TABLE I

COMPARISON OF s AND $\hat{s}$ FROM TEN ACTUAL TEST SCORE DISTRIBUTIONS

Distribution	N	s	$\hat{s}$	Error $\frac{\hat{s}-s}{s}$
1	32	3.71	3.67	-.01
2	32	6.77	6.21	-.08
3	36	9.50	9.72	+.02
4	64	12.90	12.87	-.002
5	32	8.28	8.21	-.008
6	100	12.00	12.04	+.003
7	89	1.10	1.07	-.03
8	82	5.20	5.19	-.002
9	32	8.45	7.98	-.06
10	72	3.22	3.17	-.02

FIGURE 1

ILLUSTRATING THE RELATIVE INSENSITIVITY OF $\hat{s}$ TO DEVIATIONS AWAY FROM NORMALITY



score form. Since the normal distribution is symetrical, for equal values of p, the expressions for ΣX_u and ΣX_l differ only in the sign of the average standard score term.

Returning to equation (1) and simplifying we find:

$$\hat{s} = \frac{K}{N} [2Nps (\frac{y}{p})] \qquad (4)$$

or,

$$\hat{s} = 2Kys \ .$$

Since our intention is for $\hat{s} = s$ in a normal distribution, we find:

$$2Ky = 1$$

or,

$$Ky = .500 \ .$$

Because y is a function of p, the proposition can be restated in the following way: if p is selected as some value $0 < p \leq .5$, K will be the quantity $.500/y$.

The choice of a value for p is a matter of convenience in any practical problem and usually involves choosing enough values from the extreme scores to smooth out chance fluctuations while not taking so many values that the work becomes prohibitive. One point which the author has found convenient is to set p at .167. When p is one-sixth, $y = .250$ and $K = 2$. Thus, $\hat{s}$, the approximation of s, can be expressed as: the difference between the sums of the upper and lower one-sixths of a normal distribution divided by one-half the number of scores in the distribution.

An Empirical Comparison of $\hat{s}$ and s

In the case of a normal distribution, $\hat{s}$ is identical with s. However, even when the distribution is not normal, $\hat{s}$ is still a good approximation to s. For comparative purposes a number of actual test distributions were examined, with s computed in the usual way

$$\left[s = \sqrt{\frac{\Sigma(X - \bar{X})^2}{N}} \right]$$

and then approximated by the short-cut formula proposed above. Table I presents the results of these comparisons.

In no case was the error over eight percent and in most cases in the order of two percent or less. Although the comparisons for only ten distributions are presented, they illustrate the order of differences between $\hat{s}$ and s for a great many other comparisons which the author has made.

To illustrate the relative insensitivity of $\hat{s}$ to deviations away from normality, three hypothetical distributions were constructed: 1 rectilinear; 2 bimodal; 3 highly skewed (see Figure 1).

Distributions 1 and 2 had errors of -.04. Only in the extremely skewed distribution (error -.13) was $\hat{s}$ a questionable approximation of s.

Conclusions

In all cases where actual test score distributions have been considered, $\hat{s}$ has been entirely satisfactory as an approximation of s. In computing standard scores, for example, the usual error is in the order of two to three percent. Because such an error is insignificant in all but the most demanding situations, the writer has found the approximation described here extremely appealing to teachers who consider the more conventional approaches to computing s too complicated and/or too time consuming to warrant its calculation.

REFERENCES

1. Diederich, P. B. "Simplified Measurement Techniques for Teachers," in Edith M. Hudleston (Ed.) Yearbook of the National Council on Measurements Used in Education (New York: National Council on Measurements Used in Education, 1958).
2. Wert, J. E., and others. Statistical Methods in Educational and Psychological Research (New York: Appleton-Century-Crofts, 1954).

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 3, March 1961)

COMPARING TWO METHODS OF WEIGHTING A SET OF SCORES

CHARLES E. HALL
Walter V. Clarke Associates, Inc., East Providence, R.I.

ONE OF THE more common problems of applied psychometrics is the problem of combining a collection of test or item scores into a single score. Often the psychometrist must choose between the simple sum of scores and a weighted sum, where the weights are derived according to some optimum procedure, say scaling or regression weights. In order to make a rational decision, the psychometrist must either compute the composite scores for each subject and correlate the two composites or follow a procedure like that recommended by Gulliksen (1950, p. 316 et seq.). In the author's experience, the psychometrist does not bother to go through the long, laborious computations, but rather selects from his own preferences, the weighting system to be used.

Development

Consider x_{ik}, a set of scores on k subjects over i measures. Let $l = (l_i)$ be one set of weights which might be applied to these measures to obtain a composite score as follows:

$$l x = l_1 x_{1k} + l_2 x_{2k} + \dots l_n x_{nk}.$$

Suppose $m = (m_i)$ is another such set of weights. The correlation between lx and mx can be written as follows:

$$r_{lx,\,mx} = \frac{\Sigma_k\, lx \cdot mx \; - \; (1/n)\Sigma_k lx \cdot \Sigma_k mx}{\sqrt{\Sigma_k (lx)^2 - (1/n)\,(\Sigma_k (mx)^2}\;\sqrt{\Sigma_k (mx)^2 - (1/n)\,(\Sigma_k mx)^2}}\,. \quad (1)$$

Suppose we restrict the x scores so that

$$\Sigma_k x_k^2 = 1 \text{ and } \Sigma_k x_k = 0. \quad (2)$$

With this restriction a simplification occurs.

First,
$$\begin{aligned} \Sigma_k lx &= \Sigma_k \Sigma_i l_i x_{ik} \\ &= \Sigma_i l_i (\Sigma_k x_{ik}) \\ &= \Sigma_i l_i \cdot 0 = 0, \text{ since} \end{aligned}$$
$$\Sigma_k x_{ik} = 0.$$

Equation (1) then reduces to

$$r_{lx,\,mx} = \frac{\Sigma_k lx \cdot mx}{\sqrt{\Sigma_k (lx)^2} \cdot \sqrt{\Sigma_k (mx)^2}}\,. \quad (3)$$

Now let us consider the numerator of equation (3).

$$\begin{aligned} \Sigma_k lx \cdot mx &= \Sigma_k [\,(\Sigma_i l_i x_{ik}) \cdot (\Sigma_j m_j x_{jk})] \\ &= \Sigma_k [\,\Sigma_i \Sigma_j l_j m_j x_{ik} x_{jk}] \\ &= \Sigma_i \Sigma_j l_i m_j (\Sigma_k x_{ik} x_{jk}). \end{aligned}$$

With the restrictions (2) above, $\Sigma_k x_{ik} x_{jk} = r_{ij}$; and the numerator reduces to $\Sigma_i \Sigma_j l_i m_j r_{ij} = lRm$; where R is the correlation matrix between the unweighted items or test scores and l and m are treated as vectors.

The denominator of equation (3) can be reduced by a similar process to $\sqrt{lRl'} \cdot \sqrt{mRm'}$.

Equation (1) finally reduces to

$$r_{lx,\,mx} = \frac{lRm'}{\sqrt{lRl'}\;\;\sqrt{mRm'}} \quad (4)$$

Discussion

The correlation gained from this procedure is somewhat analogous to a reliability coefficient and, to the author's thinking, ought to be handled in the same way. Although the sampling error of $r_{lx,\,mx}$ is that of the product moment correlation, the interest here is the replication of information for the two weighting systems, not the testing of a null hypothesis. Because of this, only high values of $r_{lx,\,mx}$ ought to be considered as acceptable.

It is possible to test the "un-correlation" of two sets of weights by testing the significance of the coincident coefficient of alienation,

$$a = \sqrt{1 - r^2_{lx,\,mx}}.$$

Two conveniences occur when using this technique to handle data. 1) This procedure is insensitive to multiplying all the weights of l by a constant. 2) This procedure is insensitive to the means of the scores.

To show that the set of weights $al = (al_1, al_2 \dots)$ is comparable to $l = (l_1, l_2 \dots)$, we may merely substitute al for m in equation (4).

$$\begin{aligned} r_{lx,\,alx} &= \frac{lR(la)'}{\sqrt{lRl} \cdot \sqrt{(al)R(la)}} \\ &= \frac{alRl'}{a\sqrt{(lRl\;)^2}} = 1. \end{aligned}$$

To show that the procedure is insensitive to means is more complex. First, consider the (weighted) covariance of $l(x+a)$ and lx.

TABLE I

Test Title	Guilford-Zimmerman Perceptual Speed	Guilford-Zimmerman Spatial Orientation	Guilford-Zimmerman Spatial Visualization	Thurstone's Space (Flags, Figures, Cards)	Thurstone's Form Board	Thurstone's Identical Forms	Wrightstone-O'Toole Spatial Relationships
Rotated Orthogonal Factor Coefficients							
Visualization	160	519	738	511	714	401	637
Perceptual Speed	377	269	062	075	159	515	-014
Space	333	217	-101	514	199	088	-110
Factor Score Coefficients							
Visualization	-045	118	355	132	267	038	280
Perceptual Speed	266	165	-065	-143	-005	459	-131
Space	194	116	-259	515	099	-067	-217
Standard Deviations							
	11.80	9.43	11.06	19.69	5.74	6.78	2.20
Approximate Univariate Coefficients							
Visualization	0	0	11.06	0	11.48	0	11.00
Perceptual Speed	11.80	0	0	0	0	13.56	0
Space	11.80	0	0	39.38	0	0	0
Raw Score Coefficients							
Visualization	0	0	1	0	2	0	5
Perceptual Speed	1	0	0	0	0	2	0
Space	1	0	0	2	0	0	0

TABLE II

RELATIONSHIPS BETWEEN THE TWO SCORES

Factor	r	$a = \sqrt{1 - r^2}$
Visualization	.98	.20
Perceptual Speed	.91	.41
Space	.84	.54

$\Sigma_k l(x+a)lx - (1/n)\Sigma_k l(x+a) \cdot \Sigma_k lx =$
$\Sigma(lx \cdot lx + la \cdot lx) - (1/n)[\Sigma lx + \Sigma la] \cdot \Sigma lx =$
$\Sigma lx \cdot lx + \Sigma lalx - (1/n)(\Sigma lx)^2 - (1/n)\Sigma la \Sigma lx.$

Now $la = l_1 a_1 + l_2 a_2 + \ldots l_k a_k$ is a constant; allowing the reduction
$\Sigma(lx)^2 - (1/n)(\Sigma lx)^2 + la\Sigma lx - (1/n)n \cdot la \cdot \Sigma lx =$
$\Sigma(lx)^2 - (1/n)(\Sigma lx)^2.$

Second, consider the (weighted) variance of $l(x+a)$.
$\Sigma l(x+a) \cdot l(x+a) - (1/n)[\Sigma l(x+a)]^2 =$
$\Sigma(1x)^2 + 2\Sigma lx \cdot la + \Sigma(la)^2 - (1/n)(\Sigma lx)^2$
$- (2/n)\Sigma lx \Sigma la - (1/n)(\Sigma la)^2 =$
$\Sigma(lx)^2 - (1/n)(\Sigma lx)^2 + 2la\Sigma lx - (2/n)nla\Sigma lx + n(la)^2$
$- (1/n)n^2(la)^2 = \Sigma(lx)^2 - (1/n)(\Sigma lx)^2.$

Thus the covariance of $l(x+a)$ is also equal to the variance of lx, showing that the correlation between $l(x+a)$ and lx is 1.

This technique, however, is not insensitive to the variance of the x scores.

An Example

The following example is cited to show how this technique can be used to simplify the computation of factor scores. Michael, Zimmerman and Guilford (1951) collected information about high school girls, using seven spatial relations tests. The writer factor analyzed the information using Canonical Factor Analysis and rotated three factors to a meaningful orthogonal structure. The rotated coefficients are presented in Table I. Subsequently, the factor score coefficients regressing univariate test scores onto the factor scores were computed according to the procedure outlined by Holzinger and Harman (1941, p. 268, equation 12.16). These also appear in Table I. The present objective is to simplify these regression equations by substituting simple integers or zeros for the decimals.

One of the difficulties of the procedure is that it is sensitive to the variance of the scores. Since the factor score coefficients are associated with univariate scores and it is desired to use integral coefficients with raw scores, a little juggling is required. If LX denotes the combination of integral coefficients, L, and the raw scors, X, then $(L\sigma)(\frac{X}{\sigma})$ adjusts the raw scores to univariate scores and the integral weights to comparable factor score coefficients simultaneously.

The trick to developing comparable integral weights is to find products of integers and standard deviations which are proportional to the regression coefficients. Table I also presents the standard deviations which are approximately proportioned to the factor score coefficients and the integral raw score weights. The use of these integers makes it possible to compute reasonable approximations to the factor scores with an adding machine. Table II shows the correlation between the actual factor scores and approximations obtained from integral weights. Perhaps better integral coefficients could be obtained from a more diligent search of the possibilities.

REFERENCES

Holzinger, K. J., and Harman, H. H. Factor Analysis, University of Chicago Press, 1941.

Gulliksen, H. Theory of Mental Tests, John Wiley and Sons, New York, 1950.

Michael, W. B.; Zimmerman, W. S., and Guilford, J. P. "An Investigation of the Nature of the Spatial Relations and Visualization Factors in Two High School Samples," Educational and Psychological Measurements, 1951, 11, 561-577.

The Journal of Experimental Education

PUBLISHED QUARTERLY

Copyright 1961 by
DEMBAR PUBLICATIONS, INC.
Madison 3, Wisconsin

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 4, June 1961)

A STUDY OF THE RELATIONSHIP OF SCORES ON THE MMPI TO SUCCESS IN TEACHING AS INDICATED BY SUPERVISORY RATINGS

CARROLL EDWARD FLANAGAN*
Wisconsin State College
Whitewater, Wisconsin

SECTION I

The Problem and Its Background

THE PURPOSE of this investigation is to study the relationship between certain aspects of personality and teacher efficiency. More specifically, the study is of the relationship between scores and profile patterns on the Minnesota Multiphasic Personality Inventory and success in teaching as indicated by principals' and superintendents' ratings.

There is considerable evidence from previous research to indicate that teacher ratings based on pupil gains and supervisory ratings of teachers are not highly correlated. Rostker[1]** and Lins[2], for example, each concluded that pupil gain and supervisory ratings are not related to a greater degree than can be attributed to chance. Lins stated, "Whatever is measured by each of the criteria, at least it appears evident that these criteria do not measure the same aspects of teaching efficiency."[3] LaDuke[4] summarized, "Ratings of teachers do not agree with the criterion of pupil gain."

Supervisory ratings of teachers, then, must pertain to something other than efficiency in promoting pupil gain. They probably pertain to the personal characteristics of teachers and effectiveness in discipline. An examination of the verbal statements relative to the strengths and weaknesses of the teachers rated by the principals and supervisors responding to the inquiry would seem to indicate that, except for discipline, the primary concern of the raters was with the personal characteristics of teachers.

The value systems used by principals and supervisors in rating teachers differ to a considerable degree. This, which would tend to reduce correlations to zero, is a possible reason for the low correlations between the supervisory rating and the pupil gain criteria. It would seem that any relationships which have been found between supervisory ratings and other data exist in spite of the diversity in supervisors' value systems.

Regardless of the meaning of supervisory ratings, teachers are usually judged by their supervisors and administrators. In general, these evaluations are made upon the basis of general impressions made of the teacher in many personal contacts, in observing classroom performance, and upon reports from others, rather than from a single objective measure of teaching efficiency. The impressions may depend to a large extent upon the teacher's personality, but the evaluation, whatever its basis, becomes the official evaluation of the particular school system in which the teacher is employed. Regardless of what else the teacher may have done, right or wrong, if he pleases the administration he is rated highly and consequently is considered to be a good teacher.

It can be said, also, that in spite of the inadequacies of supervisory ratings, the local administrator or supervisor should be in a better position to judge the abilities of the teacher than persons who must base their judgments on limited observation. In any case, it seems that supervisory ratings are very important when considering a teacher's success.

As to the aspect of teaching here chosen for study, a teacher's personality is apt, according to the thinking and researches of many persons, to influence his success in teaching. Various qualities affect his classroom effectiveness, his enthusiasm, his pupil contacts, his discipline, his contacts with fellow teachers and with the citizens of his community, or other aspects of his life as a teacher. Certainly, extreme abnormalities in personality determine a teacher's effectiveness in guiding pupils in the learning-teaching situation. The study here reported purports to explore the influence of certain personality abnormalities as measured by supervisory ratings.

---

* The author wishes to express his thanks to Professor A. S. Barr, who served as his major professor, and to Professors L. E. Drake and T. A. Ringness, both of whom served on the author's thesis committee.
**All footnotes will be found at end of article.

It is expected that this study will discover such relationships as may exist between teacher effectiveness as measured by supervisory ratings and scores on the MMPI.

SECTION II

The Procedures Used in the Investigation

Design of the Study

The study here reported attempts to compare the ratings of teachers who were rated outstanding, above average, average, below average, or poor by their principals or superintendents, with scores on the MMPI.

The Minnesota Multiphasic Personality Inventory

The MMPI is a psychometric instrument developed at the University of Minnesota by S. R. Hathaway, Professor and Director of the Division of Clinical Psychology at the University of Minnesota, and J. C. McKinley (deceased), formerly head of the Department of Neuropsychiatry at the University of Minnesota. The Inventory was designed to provide scores on all of the more important clinical phases of personality. Although it was standardized on 1500 normal individuals and 220 psychopathic patients in the University of Minnesota Hospital and was designed primarily as a clinical instrument, it has been used very extensively in a great variety of experiments of a nonclinical sort. Albert Ellis[5], in the Fifth Mental Measurements Yearbook[6], says, "It can confidently be stated that in the whole history of modern psychology there has been no other personality inventory on which so much theoretical and practical work has been done." The Fifth Mental Measurements Yearbook lists 495 studies about or with the MMPI. Accordingly, including those listed in previous yearbooks, there has been a total of 779 studies through 1958, relating to this particular instrument. Most of the early research on the Inventory was done in a hospital setting, but later research was conducted with normal persons and in relation to various situations. Drake and Oetting[7] state that "... almost all the published work has reported studies of the individual scales. While information about the individual scales may be useful in some instances, the behavior that the counselor is trying to understand or predict is usually too complex to be measured by a single scale." Accordingly, attention has been shifted to patterns of scales as well as individual scales.

The group form of the MMPI consists of 566 statements covering a wide range of subject matter, from physical condition of the individual being tested to his morale and attitude[8]. In administering the test, the subject is asked to respond to all statements which are in the first person, as True, False, or Don't Know. An IBM machine-scoring answer sheet has been adapted for this Inventory by Professor L. E. Drake of the University of Wisconsin Student Counseling Center, which has greatly facilitated the scoring operations.

The Inventory scales, of which there are ten plus four correction scores, were developed serially from 1942 to 1951. This probably accounts for the fact that most of the earlier research with the MMPI has been with individual scales rather than profiles.

Data-Gathering Procedures and Subjects of the Study

The MMPI was administered to entering freshmen during freshman orientation week at the University of Wisconsin. The scoring was done by the University of Wisconsin Student Counseling Center. Personality profiles for the subjects of this study were obtained from these score cards.

A list of all students who received the University Teachers' Certificate along with the teaching fields of these students was secured from the files of the Office of the Dean of the University of Wisconsin School of Education. The teaching positions held by the subjects were obtained from the files of the University Teachers' Placement Office, the University Alumni Office files, and from the official directories of the public schools of the state of Wisconsin.

Letters were sent to all known principals and superintendents of teachers in the study asking their cooperation in rating the teachers involved. The rating blank asked for comments regarding elements of strength in the teacher's work, the elements of weakness in the teacher's work, general comments, and an overall rating on a five-point scale.

Tabulations of the ratings, comments, and patterns of personality scores were then made and relationships among the various data sought.

The subjects for this study consisted of all the students enrolled at the University of Wisconsin during the fall semesters of 1949, 1950, and 1951, who received the University Teachers' Certificate in 1953, 1954, 1955, or 1956, and who have taught in the public schools of Wisconsin. It was found that a total of 431 students from the freshman classes of 1949, 1950, and 1951 had taken the MMPI, had graduated in the years 1953-56, and had received the University Teachers' Certificate. Due to the Korean situation and other matters, many of the males who had originally enrolled did not graduate and many more who did graduate did not go into teaching. Marriage and decisions not to teach removed many of the females from the list of teachers. For others, no record of post-graduation activity could be found. In some cases, due to the death of a principal or superintendent, or to his moving to an unknown location, it was impossible to get a supervisory rating on a particular teacher. In a few cases the administrators and supervisors failed to supply the ratings. The number of teachers for whom complete data were secured was 167, of which only 20 were males.

Although this is not intended to be a sampling study, there appears to be no reason to believe that the people here studied were greatly different from other people who graduate from the University of Wisconsin School of Education. Since this is a very limited population, users of the data in this study may be interested to know more about the group. Considering the group of 167 as a unit, this population was different from the usual in that only twenty were males. For this reason, and because high or low scores on many of the MMPI scales mean different things for men and women, the female population of 147 was studied separately. Of this group, nine were music teachers; twelve, teachers of art; thirteen, speech correction; fifteen, physical education and recreation; thirty-three, home economics; six, business education; fifteen, elementary; eight, social studies; one, mathematics and science; and thirty-five, English, foreigh language, speech or library. All had grade-point averages of 2.5 or better at the beginning of their junior year. In general, they appeared to be a typical University of Wisconsin School of Education group. Assuming there has not been much change in the basic characteristics of University of Wisconsin students, the interested reader is referred to Lins[9] study for detailed information concerning a typical university group of women students.

Whenever it was possible, which was in more than half the cases, two or more ratings were secured on each teacher. When these ratings did not agree, they were averaged and the average rating used. When ratings differed, they differed by only one rank.

Of the 147 women, ten were rated outstanding or 1; eight had a mean rating of 1.5; sixty-one were rated above average or 2; fourteen had a mean rating of 2.5; forty-two were rated average or 3; three had an average rating of 3.5; seven teachers were rated below average or 4; and two were rated poor or 5. Either most of the teachers in this study were above average or supervisors and principals have a tendency to overrate their teachers. Because the frequencies in some of the categories were rather small, they were combined as follows: eighteen teachers in the 1 or 1.5 rating group, seventy-five cases in the 2 or 2.5 rating category, forty-five in 3 or 3.5, and nine cases in the 4 or 5 rating group.

Coding the Data

The scores on the MMPI had been coded according to the method of Hathaway.[10] The scales were assigned numbers as follows:

0 Si - Social Introversion-Extroversion
1 Hs - Hypochondriasis
2 D - Depression
3 Hy - Hysteria
4 Pd - Psychopathic Deviate
5 Mf - Masculinity-Femininity
6 Pa - Paranoid
7 Pt - Psychasthenia
8 Sc - Schizophrenia
9 Ma - Hypomania
X None

Raw scores had first been converted to standard T scores. (T scores have a mean of 50 and a standard deviation of 10). In the coding, the number of the scale with the highest T score was written first, then the second highest, and so on down. No score which fell within half a standard deviation on either side of the mean was recorded. This means that only scores of 55 or above were coded high. After the high scores had been listed according to the number of their scales, a dash separated the low scored scales. These were listed by scale number with the lowest scored scale first, the next lowest second, and so on. No score above 45 was coded low. Tied scores were underlined. Thus, if a person's profile code is 3987-645, it means that the highest score was made on scale 3, the score on scale 9 was equal to that on 3 or differed by only one point, the next highest score was on scale 8, and the next on scale 7. Scores of 55 or above were made on all of these scales. The lowest score was made on scale 6, the next lowest on scale 4, and the next on scale 5. The scores on these three scales were all below 46. Scores on the remaining scales (0, 1, and 2) were within the normal or average range.

This system of coding has been recommended because it expresses both intensity and pattern and at the same time effects a compromise between too much detail on scale elevation and equally undesirable detail as to profile shape[11]. Du Mas[12], in discussing problems of profile analysis concludes, "If prediction or diagnosis is to be maximally good, both intensity and pattern must be operative in a profile."

Methods of Analyzing the Data

There are many theories of mental structure, organization, and techniques for the validation of data-gathering devices with reference to these.

The conventional statistical techniques for test validation appear to rest upon the assumption of the additive nature of the constituents of human abilities and personalities. This study does not make this assumption and, accordingly, a different technique of validation has been employed.

It is thought by many researchers in personality that patterns of personality traits, the ways in which various phases go together, are the important determiners of personality manifestations. In some cases one scale reinforces another, while in others one scale compensates for or offsets another. For certain scales extreme scores seem to be the only important ones. The most promising recent research, especially that of S. R. Hathaway and L. E. Drake, has been conducted with reference to these ideas.

In the present study, comparisons were made between supervisory ratings and highest ranking scales, combinations of high ranking scales, and combinations of high scales and low scales. It is hypothesized that personality profiles on the MMPI are different for highly rated teachers than for teachers who are rated lower. McCall[13] found a relatively close relationship between teacher personality and pupils' comprehensive growth as measured in his investigation.

Various procedures were employed in analyzing the data. Among these was the chi-square. When a chi-square large enough to be encouraging was found, the data were examined more closely and other tests were made to see if a relationship between a particular rating and a scale or profile pattern might be more firmly established. Since in most cases the chi-square statistic was used to test whether having a particular scale or combination coded in a certain manner or not having it coded thus was statistically significant in regard to supervisory ratings, the expected frequency of each cell becomes a matter of interest.

Inasmuch as there are forty cells in a 4 × 10 table which was employed in the overall chi-square analysis, the expected frequencies for 147 cases would naturally not be large. In many instances, they fell below the minimum of five or ten frequently thought necessary for such analyses.

Under the null hypothesis it is assumed that having a particular personality trait has no relationship to supervisory ratings. For example, as shown in Table I, twenty-four female teachers had scale 3 (Hy) coded highest of all the scales. Under the null hypothesis, one would expect these twenty-four cases to be distributed among the supervisory ratings proportionately to the way in which the total 147 cases are distributed among the supervisory ratings. Seventy-five of the 147 teachers were rated in the second category. Then the expected frequency for the cell designating scale 3 coded highest and supervisory rating of 2-2.5 would be that part of twenty-four that seventy-five is of 147. This is 12.24.

In this use of chi-square, the data are put into various categories (e. g., the dichotomy of having a scale coded high or not high on one hand and the four-fold categorization of supervisory ratings on the other). These categories should not be interpreted as points on a scale.

Chi-square is a measure of the amount that observed frequencies differ from expected frequencies in the various categories. In this study most of the observed frequencies were about what one would expect. This is indicated by the chi-square values, only a few of which were large enough to be statistically significant.

Tate and Clelland[14] say, "The test (chi-square) is the best available for enumeration data from a manifold population. It gives good results even when the expected frequency in one or more categories is as low as one, provided that the majority are about five or more." Not all authors agree with Tate and Clelland's statement regarding the use of chi-square and particularly the minimum number required for expected frequencies. Since the number in each category is rather small in some cases (after the data are broken down into sub-categories) chi-squares have been figured in several instances where the statistical significance of the result must be seriously questioned. Many authors do[15,16,17], however, indicate that the heretofore commonly accepted minimum expected frequency of five per cell is too high, especially when there are two or more degrees of freedom. In this study there are three degrees of freedom for most of the chi-square tests.

In addition to the chi-square tests, the percentages of occurrence of the various patterns of high and low scores were calculated for each of the four rating categories. It was thought that such data would enable one to make certain qualitative analyses of profile patterns in relation to the supervisory rating. The following profile patterns were compared:

1. The frequency with which each scale of the three highest scales occurred with every other scale coded high (above T = 54).
2. The frequency with which each scale of the three highest scales appeared with each of the two lowest scales. This was done for each of the four rating categories.
3. The frequency with which each scale coded anywhere high (above T = 54) occurred with each scale coded anywhere low (below T = 46). This was done for the highest rated group.

The data obtained in these comparisons are shown in the tables reproduced in Section III.

TABLE I

FREQUENCY OF SUPERVISORY RATINGS OF FEMALE TEACHERS WHO HAD VARIOUS MMPI SCALES CODED HIGHEST OR TIED FOR HIGHEST

Scale Coded Highest or Tied for Highest		Supervisory Rating				
		1 - 1.5	2 - 2.5	3 - 3.5	4 - 4	Chi Square
N =	147	18	75	45	9	D. F. = 3
0	11	2 (1. 35)*	7 (5. 61)	0 (3. 37)	2 (0. 67)	7. 177
1	3	1	1	1	0	
2	3	1	1	1	0	
3	24	7 (2. 94)	11 (12. 24)	5 (7. 35)	1 (1. 47)	7. 934
4	11	2 (1. 35)	4 (5. 61)	4 (3. 37)	1 (0. 67)	1. 142
5	18	2 (2. 20)	12 (9. 18)	4 (5. 51)	0 (1. 10)	2. 733
6	18	1 (2. 20)	10 (9. 18)	6 (5. 51)	1 (1. 10	. 893
7	30	1 (3. 67)	16 (15. 31)	11 (9. 18)	2 (1. 84)	2. 954
8	8	0	5	2	1	
9	30	4 (3. 67)	16 (15. 31)	9 (9. 18)	1 (1. 84)	. 560
X	4	1	0	3	0	

*Number in parentheses following observed frequency is expected frequency for that cell.

SECTION III

The Analysis of Data

The purpose of this study, as stated in Section I, is to explore the relationship between scores and profile patterns on the MMPI and success in teaching as indicated by supervisory ratings. The study is designed to discover whether or not there are differences in the personality patterns of successful and unsuccessful teachers.

In order to carry out the stated purposes of this study, it was necessary to submit the data to certain statistical analyses which will be explained. The data in this study are nonparametric in nature and, consequently, any statistical measures employed should be of the nonparametric type. The methods of securing the supervisory ratings and of grouping them into four categories were explained in Section II, as was the method of coding the scores from the MMPI.

The Female Teachers

The following discussion is in reference to the group of 147 female teachers only.

The frequency with which each scale was coded highest or tied for highest for each of the four supervisory rating categories was first tabulated (Table I). Because of the low frequencies in some of the categories certain planned techniques had to be abandoned. However, chi-squares were figured for cases which had a total N of 11 or more. Whenever the chi-square value was large enough to suggest some relationship, closer examination of the data were made. In most cases chi-square values were calculated using a method advocated by Lindquist[18] which is less time consuming than the traditional method, but which is the algebraic equivalent of the traditional.[19]

Only one group—the group with scale 3 (Hy) coded highest—seemed worthy of further investigation. A chi-square of 7.934 with three degrees of freedom was found for this group, indicating that having scale 3 (Hy) coded highest might discriminate among the female teachers in the various rating categories. With three degrees of freedom, a chi-square of 7.934 is significant beyond the 5 percent level.

For the group with scale 3 coded highest, the expected frequencies for each rating, under the null hypothesis of no difference, were 2.94, 12.24, 7.35, and 1.47 for ratings of 1 - 1.5, 2 - 2.5, 3 - 3.5, and 4 - 5, respectively. A comparison of the expected frequencies with the observed frequencies indicated that the greatest discrepancy was for the 1 - 1.5 rating. It was then hypothesized that being rated high, i.e., 1 or 1.5 was related to having scale 3 coded highest of all the scales. The data were arranged in a 2 × 2 contingency table, Yates' correction[20] for small expected frequencies applied, and chi-square calculated. The value found was 5.871 (see Table IX). With one degree of freedom, this value of chi-square is significant beyond the 2 percent level. Examination of the data showed the relationship to be positive.

For these data at least, there seems to be a potential relationship for female teachers between the occurrence of scale 3 (Hy) being coded highest and a supervisory rating of outstanding. Much larger groups of teachers including many who have been rated outstanding are needed to establish this relationship firmly.

Continued exploration of various relationships among the data tabulated as indicated above failed to reveal additional statistical significant relationships. However, an examination of the relative frequencies in the various categories did bring out some interesting relationships which will be pointed out. Although only eleven cases out of the total 147 female teachers had scale 0 (Si) coded highest, only two of these received supervisory ratings of average or less. It would appear that an absence of scale 0 (Si) coded highest may be associated with average or below average supervisory ratings.

However, when one considers the small number (eleven) in the group, it is apparent that this may be a false conclusion. When the group is enlarged to include all those who had scale 0 (Si) coded among the three highest scales (see Table II), the number is increased to thirty-one and the number receiving ratings of average or below is increased to thirteen, which is more than the expected frequency. This would seem to dispute the conclusion of relationship between a lack of scale 0 (Si) being coded high and average or below average supervisory ratings.

The above is an example of a fact that has been obvious throughout the analysis of the data in this study, namely, that for studies of this nature very large groups, preferably in the thousands rather than the hundreds, are highly desirable. Whereas normally a much smaller group is ample for a good investigation, studies of this kind necessitate a group large enough to provide sufficient cases in each of the sub-groups after the data have been broken down.

A similar examination of the group with scale 7 (Pt) coded highest indicates a possible slight relationship between scale 7 (Pt) coded highest and average or below supervisory ratings. When the relationship between scale 7 (Pt) being coded among the three highest and supervisory ratings is investigated, support is gained for this possibility, since the chi-square for this relationship is 5.254 (about the 16 percent level). Although the evidence is not statistically significant at the commonly designated 5 percent level, analysis of the data for this study seems to give some indication that high codings of scale 7 (Pt) are not

TABLE II

FREQUENCY OF SUPERVISORY RATINGS OF FEMALE TEACHERS WHO HAD VARIOUS MMPI SCALES CODED AMONG THE THREE HIGHEST

Scale Coded Among the Three Highest		Supervisory Rating				Chi-Square
		1 - 1.5	2 - 2.5	3 - 3.5	4 - 5	
N =	147	18	75	45	9	D. F. = 3
0	31	4 (3.80)*	14 (15.82)	11 (9.49)	2 (1.90)	
1	14	2 (1.71)	8 (7.14)	4 (4.29)	0 (0.86)	1.135
2	16	1 (1.96)	9 (8.16)	4 (4.90)	2 (0.98)	2.178
3	54	9 (6.61)	25 (27.55)	15 (16.53)	5 (3.31)	.774
4	45	5 (5.51)	25 (22.96)	13 (13.78)	2 (2.76)	.691
5	29	3 (3.55)	20 (14.80)	6 (8.88)	0 (1.78)	5.761
6	45	6 (5.51)	22 (22.96)	15 (13.78)	2 (2.76)	.576
7	62	4 (7.59)	32 (31.63)	20 (18.98)	6 (3.80)	5.254
8	42	3 (5.14)	24 (21.43)	12 (12.86)	3 (2.57)	1.862
9	48	6 (5.88)	25 (24.49)	15 (14.69)	2 (2.94)	.474
X	4	1	0	3	0	

*Number in parentheses following observed frequency is expected frequency for that cell.

TABLE III

FREQUENCY OF SUPERVISORY RATINGS OF FEMALE TEACHERS WHO HAD VARIOUS MMPI SCALES CODED LOWEST OR TIED FOR LOWEST

Scale Coded Lowest or Tied for Lowest		Supervisory Rating 1 - 1.5	2 - 2.5	3 - 3.5	4 - 5	Chi-Square
N =	147	18	75	45	9	D. F. = 3
0	21	3 (2. 57)*	13 (10. 71)	3 (6. 43)	2 (1. 29)	3. 248
1	5	0	3	2	0	
2	16	3 (1. 96)	10 (8. 16)	3 (4. 90)	0 (0. 98)	3. 009
3	7	1	5	1	0	
4	4	1	2	0	1	
5	41	4 (5. 02)	17 (20. 92)	17 (12. 55)	3 (2. 51)	3. 625
6	7	0	3	4	0	
7	13	2 (1. 59)	3 (6. 63)	8 (3. 98)	0 (0. 80)	7. 626
8	3	1	1	1	0	
9	11	1 (1. 35)	7 (5. 61)	1 (3. 37)	2 (0. 67)	5. 091
X	19	2 (2. 33)	11 (9. 69)	5 (5. 82)	1 (1. 16)	. 413

*Number in parentheses following observed frequency is expected frequency for that cell.

associated with high supervisory ratings for women teachers, but that low codings of scale 7 are associated with average or below supervisory ratings for women teachers (see Tables III, IV, VII and IX). The code type 7 - 5 (high coding of seven with low coding of 5) was a common one for women teachers in both the above average and average or below rated groups, the evidence again favoring the lower rated teachers. Superior teachers did not, in general, make their highest MMPI scores on scale 7 (see Tables I and II).

Further examination of the data revealed a possible meaningful relationship between scale 5 (Mf) coded among the three highest and supervisory ratings. The chi-square value indicating the relationship was 5. 761 (about the 13 percent level). This was high enough to merit further investigation (see Table II).

The number in the group which had scale 5 (Mf) coded among the three highest was 29. The expected frequencies for the four supervisory ratings from 1 to 5 were, respectively, 3. 55, 14. 80, 8. 88, and 1. 78. The largest discrepancy between expected and observed frequencies was for the group with ratings of 2 or 2. 5. When the chi-square test was applied to the re-grouped data to test whether scale 5 being coded among the three highest might be related to a 2 or 2. 5 rating, a value of 4. 65 was obtained. If Yates' correction for small expected frequencies is applied, the calculated chi-square value is 3.80 (see Table IX). With one degree of freedom, 3.80 just misses being significant at the 5 percent level.

The data for this same group were then arranged in a 2 × 2 contingency table to test whether having scale 5 (Mf) coded among the three highest scores might be related to ratings anywhere above average, that is, from 1 through 2. 5. With this arrangement, Yates' correction was unnecessary since the expected frequencies in each cell were above ten. This time a chi-square of 4. 00 was found (see Table VII). This is significant beyond the 5 percent level, indicating that for the group of women teachers in this study, a high coding of scale 5 is positively related to above average or outstanding supervisory ratings. There is some indication that the rating is more apt to be 2 or 2. 5 than 1 or 1. 5.

There is some indication that a low coding of scale 5 (Mf) for the women teachers is associated with average or below average supervisory ratings (see Tables III, IV, VII, and VIII).

A comparison of the frequencies of various combinations of high and low coded scales for the "above" and "below" rated teachers did not indicate any statistically significant differences. The number in most categories was too small to be dependable (see Table VIII). An examination of the data, however, shows that for every code type with 5 low (0 - 5, 1 - 5, 2 - 5, 3 - 5, etc.,) the expected frequency is larger than the observed frequency for the average or below average group.

Since scale 5 coded low has been found to be either a suppressor or an intensifier factor with high codings of other scales[21], chi-square tests were run to see if a low coding of scale 5 with a high coding of each of the other scales differentiated between supervisory ratings above average and those below that. For the group of women teachers studied, no sigificant differences in supervisory ratings were found within the group with a particular scale coded high in respect to having scale 5 coded low or not coded low. Data for this analysis can be found in Table VI.

A low coding of scale 2 (D) was associated with the group of women teachers studied with above average or outstanding supervisory ratings (see Tables III, IV, and VII). When the supervisory ratings were divided into a dichotomy, a chi-square value at the 5 percent level was calculated (see Table VII). Drake[22] says that a low coding of scale 2 for women is frequently related to "socially outgoing characteristics".

The proportion of the number in each supervisory rating category who had the various scales coded highest or among the highest was also calculated and recorded as a percent (see Appendix B, Tables A and B). * Tabulations of other combinations were also made (see Appendix B, Tables C, D, and E). While some of the relationships shown in these tables may be of some interest, none of these supplementary tables indicate anything of statistical significance.

The Male Teachers

The number of male teachers in the group studied was so small (twenty) that little could be done of a statistical nature. When the available data were broken down into the necessary subgroups, the number in each group was usually zero or one and seldom more than two. Descriptive statements concerning this group of male teachers will be made as far as seems appropriate.

Two of the twenty men in the group received supervisory ratings of 1. 5; thirteen of the group were rated 2; one was rated 2. 5, and four were rated 3. Thus, sixteen of the twenty men in the group were rated above average or outstanding by their supervisors or principals. The other four were rated average. With numbers this small, any statements of relationships between supervisory ratings and MMPI code types could not be

---

*All references to Appendices will be found in original thesis on file, Library, University of Wisconsin, Madison, Wisconsin.

TABLE IV

FREQUENCY OF SUPERVISORY RATINGS OF FEMALE TEACHERS WHO HAD VARIOUS MMPI SCALES CODED AMONG THE TWO LOWEST

Scale Coded Among the Two Lowest		Supervisory Rating				Chi-Square
		1 - 1.5	2 - 2.5	3 - 3.5	4 - 5	
N	= 147	18	75	45	9	D. F. = 3
0	36	4 (4.41)	20 (18.37)	9 (11.02)	3 (2.20)	1.113
1	22	3 (2.69)	12 (11.22)	6 (6.73)	1 (1.35)	.303
2	32	5 (3.92)	20 (16.33)	6 (9.80)	1 (1.96)	3.919
3	14	1 (1.71)	8 (7.14)	5 (4.29)	0 (0.86)	1.522
4	16	2 (1.96)	7 (8.16)	6 (4.90)	1 (0.98)	.466
5	50	5 (6.12)	22 (25.51)	18 (15.31)	5 (3.06)	3.623
6	15	2 (1.84)	7 (7.65)	6 (4.59)	0 (0.92)	1.582
7	27	5 (3.31)	11 (13.78)	11 (8.27)	0 (1.65)	4.882
8	10	2 (1.22)	4 (5.10)	4 (3.06)	0 (0.61)	1.764
9	23	4 (2.82)	13 (11.73)	4 (7.04)	2 (1.41)	2.603
X	19	2 (2.33)	11 (9.69)	5 (5.82)	1 (1.16)	.413

*Number in parentheses following observed frequency is expected frequency for that cell.

TABLE V

FREQUENCY OF SUPERVISORY RATINGS OF FEMALE TEACHERS WHO HAD VARIOUS PAIRS OF MMPI SCALES CODED AMONG THE THREE HIGHEST

Pairs of Scales Among the Three Highest		Supervisory Rating 1 - 1.5	2 - 2.5	3 - 3.5	4 - 5	Chi-Square
N	= 147	18	75	45	9	D. F. = 3
01	4	1	1	2	0	
02	5	1	1	2	1	
03	7	1	4	1	1	
04	4	1	3	0	0	
05	3	1	1	1	0	
06	8	1	3	4	0	
07	21	2 (2.57)*	9 (10.71)	8 (6.43)	2 (1.29)	1.379
08	11	0 (1.35)	7 (5.61)	4 (3.37)	0 (0.67)	2.683
09	2	0	0	2	0	
0X	0	0	0	0	0	
12	1	0	1	0	0	
13	8	2	4	2	0	
14	3	0	1	2	0	
15	1	0	1	0	0	
16	2	0	2	0	0	
17	4	0	3	1	0	
18	1	0	1	0	0	
19	5	1	3	1	0	
IX	0	0	0	0	0	
23	4	0	3	1	0	
24	5	1	4	0	0	
25	1	0	1	0	0	
26	2	0	1	1	0	
27	10	0	5	3	2	
28	5	0	3	1	1	
29	1	0	0	1	0	
2X	0	0	0	0	0	
34	15	2 (1.84)	6 (7.65)	5 (4.59)	2 (0.92)	1.873
35	9	1	6	2	0	

TABLE V (Continued)

Pairs of Scales Among the Three Highest		Supervisory Rating				Chi-Square
		1 - 1.5	2 - 2.5	3 - 3.5	4 - 5	
N =	147	18	75	45	9	D. F. = 3
36	19	2 (2.33)	10 (9.69)	6 (5.82)	1 (1.16)	.097
37	11	0 (1.35)	6 (5.61)	2 (3.37)	3 (0.67)	10.772
38	6	0	3	2	1	
39	15	3 (1.84)	6 (7.65)	5 (4.59)	1 (0.92)	1.267
3X	5	1	2	2	0	
45	5	0	4	1	0	
46	11	2 (1.35)	5 (5.61)	4 (3.37)	0 (0.67)	1.271
47	20	1 (2.45)	11 (10.20)	7 (6.12)	1 (1.22)	1.257
48	10	0	6	4	0	
49	16	2 (1.96)	10 (8.16)	4 (4.90)	0 (0.98)	1.749
4X	0	0	0	0	0	
56	6	0	4	2	0	
57	4	1	1	2	0	
58	2	0	2	0	0	
59	11	0 (1.35)	9 (5.61)	2 (3.37)	0 (0.67)	.499
5X	5	1	3	1	0	
67	19	2 (2.33)	9 (9.69)	8 (5.82)	0 (1.16)	2.387
68	11	2 (1.35)	5 (5.61)	4 (3.37)	0 (0.67)	1.271
69	8	2	2	3	1	
6X	2	0	1	0	1	
78	31	2 (3.80)	17 (15.82)	9 (9.49)	3 (1.90)	2.032
79	15	1 (1.84)	9 (7.65)	4 (4.59)	1 (0.92)	.782
7X	0	0	0	0	0	
89	13	2 (1.59)	7 (6.63)	3 (3.98)	1 (0.80)	.459
8X	0	0	0	0	0	
9X	4	0	1	3	0	

*Number in parentheses following observed frequency is expected frequency for that cell.

TABLE VI

FREQUENCIES OF PRESENCE OR ABSENCE OF LOW CODING OF SCALE 5(Mf) WITH HIGH CODING OF OTHER SCALES FOR FEMALE TEACHERS WITH ABOVE AVERAGE OR SUPERIOR AND AVERAGE OR BELOW SUPERVISORY RATINGS

	Supervisory Rating		
Code Type* High - Low	1 - 2.5 N = 93	3 - 5 N = 54	Chi-Square D. F. = 1
0 - 5	6	7	
0 - Not 5	12	6	1.304
3 - 5	8	7	
3 - Not 5	26	13	.826
4 - 5	10	8	
4 - Not 5	20	7	1.667
6 - 5	12	7	
6 - Not 5	16	10	.001
7 - 5	16	15	
7 - Not 5	20	11	1.060
8 - 5	10	8	
8 - Not 5	17	7	1.046
9 - 5	7	7	
9 - Not 5	24	10	1.838

*High coded scale was among the three highest; 5, when coded low was among the two lowest.

TABLE VII

COMPARISON OF FREQUENCY OF OCCURRENCE OF CERTAIN CODE TYPES FOR FEMALE TEACHERS WHO WERE RATED ABOVE AVERAGE OR SUPERIOR WITH THOSE WHO WERE RATED AVERAGE OR BELOW AVERAGE

Code Type	Supervisory Rating 1 - 2.5 N = 93	Supervisory Rating 3 - 5 N = 54	Chi-Square D. F. = 1
0 highest	9 (6.96)*	2 (4.04)	1.760
3	18 (15.18)	6 (8.82)	1.704
4	6 (6.96)	5 (4.04)	.556
5	14 (11.39)	4 (6.61)	1.869
6	11 (11.39)	7 (6.61)	
7	17 (18.98)	13 (11.02)	.707
9	20 (18.98)	13 (11.02)	.188
0 among 3 highest	18 (19.61)	13 (11.39)	.460
1	10 (8.86)	4 (5.14)	.442
2	10 (10.12)	6 (5.88)	
3	34 (34.16)	20 (19.84)	
4	30 (28.47)	15 (16.53)	.323
5	23 (18.35)	6 (10.65)	4.002
6	28 (28.47)	17 (16.53)	
7	36 (39.22)	26 (22.78)	1.245
8	27 (26.57)	15 (15.43)	
9	31 (30.37)	17 (17.63)	
0 lowest	16 (13.28)	5 (7.72)	1.761
2	13 (10.12)	3 (5.88)	2.499
5	21 (25.94)	20 (15.06)	3.550
7	5 (8.22)	8 (4.78)	3.775**
9	8 (6.96)	3 (4.04)	.458
2 among 2 lowest	25 (20.25)	7 (11.75)	3.886
5	27 (31.63)	23 (18.37)	2.799
7	16 (17.09)	11 (9.91)	.228
9	17 (14.55)	6 (8.45)	1.330
07 among 3 highest	11 (13.28)	10 (7.72)	1.249
34 among 3 highest	8 (9.49)	7 (5.51)	.709
37	6 (6.96)	5 (4.04)	.389
49	12 (10.12)	4 (5.88)	1.066
67	11 (12.02)	8 (6.98)	.271
78	19 (19.62)	12 (11.38)	.066
89	9 (8.23)	4 (4.77)	.215

* Numbers following observed frequencies are expected frequencies.

**Chi-square for 7 coded lowest was reduced to 2.715 when Yates' correction was applied.

TABLE VIII

COMPARISON OF FEMALE TEACHERS RATED ABOVE AVERAGE OR SUPERIOR WITH THOSE RATED AVERAGE OR BELOW AS TO COMBINATIONS OF HIGH AND LOW CODED SCALES

Code Type High with Low	Above Frequency Observed	Above Frequency (Expected)	Below Frequency Observed	Below Frequency (Expected)	Chi-Square D. F. = 1
0 - 1	4	(3.80)	2	(2.20)	
0 - 2	3		1		
0 - 3	5	(5.06)	3	(2.94)	
0 - 4	1		1		
0 - 5	6	(8.23)	7	(4.77)	1.8064
0 - 6	2		2		
0 - 7	0		1		
0 - 8	0		0		
0 - 9	7	(6.96)	4	(4.04)	
0 - X	3	(2.53)	1	(1.47)	
1 - 0	1		0		
1 - 2	3		1		
1 - 3	0		1		
1 - 4	2		0		
1 - 5	2	(2.53)	2	(1.47)	
1 - 6	1		0		
1 - 7	1		0		
1 - 8	0		0		
1 - 9	3		1		
1 - X	2		1		
2 - 0	2		0		
2 - 1	0		1		
2 - 3	0		1		
2 - 4	0		0		
2 - 5	3	(4.43)	4	(2.57)	
2 - 6	1		0		
2 - 7	0		0		
2 - 8	1		0		
2 - 9	1		2		
2 - X	3		0		
3 - 0	11	(12.65)	9	(7.35)	.6779
3 - 1	4	(3.16)	1	(1.84)	
3 - 2	7	(5.06)	1	(2.94)	
3 - 4	3		2		
3 - 5	8	(9.49)	7	(5.51)	.7092
3 - 6	2		0		
3 - 7	7	(8.23)	6	(4.77)	.5495
3 - 8	5	(4.43)	2	(2.57)	
3 - 9	7	(6.33)	3	(3.67)	
3 - X	5	(5.06)	3	(2.94)	
4 - 0	7	(6.96)	4	(4.04)	
4 - 1	1		0		
4 - 2	5	(3.80)	1	(2.20)	
4 - 3	4	(3.16)	1	(1.84)	
4 - 5	10	(11.39)	8	(6.61)	.5264
4 - 6	2		1		
4 - 7	5		1		
4 - 8	0		2		
4 - 9	3		1		
4 - X	6	(5.06)	2	(2.94)	
5 - 0	8	(6.33)	2	(3.67)	.2881
5 - 1	7	(5.06)	1	(2.94)	
5 - 2	7	(6.96)	4	(4.04)	
5 - 3	2		0		
5 - 4	4		2		
5 - 6	5		2		
		(6.96)	3	(4.04)	.4561

TABLE VIII (Continued)

Code Type High with Low	Above Frequency Observed	Above Frequency (Expected)	Below Frequency Observed	Below Frequency (Expected)	Chi-Square D. F. = 1
5 - 8	2		0		
5 - 9	5		0		
5 - X	1		0		
6 - 0	7	(5.69)	2	(3.31)	
6 - 1	4		1		
6 - 2	7	(5.06)	1	(2.94)	
6 - 3	2		0		
6 - 4	4		4		
6 - 5	12	(12.02)	7	(6.98)	
6 - 7	5		3		
6 - 8	2		1		
6 - 9	8	(6.96)	3	(4.04)	.4561
6 - X	2		2		
7 - 0	3		2		
7 - 1	3		2		
7 - 2	8	(6.33)	2	(3.67)	1.2881
7 - 3	6		3		
7 - 4	2		0		
7 - 5	16	(19.61)	15	(11.39)	2.2922
7 - 6	1		3		
7 - 8	0		1		
7 - 9	6	(5.69)	3	(3.31)	
7 - X	6	(6.33)	4	(3.67)	
8 - 0	4		1		
8 - 1	2		0		
8 - 2	6	(4.43)	1	(2.57)	
8 - 3	3		2		
8 - 4	0		0		
8 - 5	10	(11.39)	8	(6.61)	.5264
8 - 6	1		1		
8 - 7	1		0		
8 - 9	2		1		
8 - X	5	(5.06)	3	(2.94)	
9 - 0	13	(11.39)	5	(6.61)	.5264
9 - 1	7	(6.96)	4	(4.04)	
9 - 2	10	(8.86)	4	(5.14)	.4415
9 - 3	1		1		
9 - 4	2		1		
9 - 5	7	(8.86)	7	(5.14)	1.1755
9 - 6	2		1		
9 - 7	7	(8.86)	7	(5.14)	1.1755
9 - 8	1		1		
9 - X	6	(5.06)	2	(2.94)	
X - 0	1		0		
X - 1	0		1		
X - 2	1		0		
X - 3	0		1		
X - 4	0		2		
X - 5	0		0		
X - 6	0		2		
X - 7	1		0		
X - 8	0		1		
X - 9	0		0		

Note: Number rated above average - 93 (63.27%)
Number rated average or below - 54 (36.73%)

High coded scale was among the three highest.
Low coded scale was among the two lowest.

TABLE IX

FREQUENCY RELATIONSHIPS AMONG MISCELLANEOUS CODE TYPES AND SUPERVISORY RATING CATEGORIES FOR FEMALE TEACHERS

	Rating		Chi-Square
Code Type	1 - 1.5	Below 1.5	D. F. = 1
3 highest	7	17	7. 643 5. 871 with Yates' Correction
3 not highest	11	112	
7 highest	1	29	2. 786
7 not highest	17	100	
3 among 3 highest	9	45	1. 556
3 not among highest	9	84	
7 among 3 highest	4	58	3. 349 2. 479 with Yates' Correction
7 not among highest	14	71	
8 among 3 highest	3	39	1. 424
8 not among highest	15	90	
47 among 3 highest	1	19	1. 131
47 not among highest	17	110	
78 among 3 highest	2	29	1. 227
78 not among highest	16	100	
7 among 2 lowest	5	22	1. 211
7 not among lowest	13	107	
	2 - 2. 5	Not 2 - 2. 5	
5 among 3 highest	20	9	4. 648 3. 80 with Yates' Correction
5 not among highest	55	63	
	3 - 3. 5	Not 3 - 3. 5	
5 lowest	17	24	3. 152
5 not lowest	28	78	
	4 - 5	Not 4 - 5	
7 among 3 highest	6	56	2. 349
7 not among highest	3	82	
7 among 2 lowest	0	27	2. 157
7 not among lowest	9	111	

TABLE X

FREQUENCY OF SUPERVISORY RATINGS OF MALE TEACHERS WHO HAD VARIOUS MMPI SCALES CODED HIGHEST OR TIED FOR HIGHEST

Scale Coded Highest or Tied for Highest	Supervisory Rating				Total N = 20
	1.5 N = 2	2. N = 13	2.5 N = 1	3. N = 4	
0	0	0	0	0	0
1	0	1	0	0	1
2	0	1	0	0	1
3	0	2	0	0	2
4	0	0	0	1	1
5	2	5	0	2	9
6	1	2	0	0	3
7	1	2	0	1	4
8	0	0	0	0	0
9	0	1	1	0	2

TABLE XI

FREQUENCY OF SUPERVISORY RATINGS OF MALE TEACHERS WHO HAD VARIOUS MMPI SCALES CODED AMONG THE THREE HIGHEST

Scale Coded Among Three Highest	Supervisory Rating				Total
	1.5	2.	2.5	3.	
0	0	2	0	0	2
1	0	2	1	0	3
2	0	3	0	2	5
3	1	5	1	0	7
4	0	2	0	2	4
5	2	7	2	2	13
6	1	4	0	0	5
7	1	3	0	2	6
8	1	4	1	1	7
9	1	1	1	1	4
X	0	0	0	0	0

TABLE XII

FREQUENCY OF SUPERVISORY RATINGS OF MALE TEACHERS WHO HAD VARIOUS MMPI SCALES CODED LOWEST OR TIED FOR LOWEST

Scale Coded Lowest or Tied for Lowest	Supervisory Rating				Total
	1. 5	2	2. 5	3	
0	1	2	0	0	3
1	0	1	0	1	2
2	0	0	0	1	1
3	0	1	0	0	1
4	0	1	0	0	1
5	0	0	0	1	1
6	0	1	0	0	1
7	0	2	0	0	2
8	0	2	0	0	2
9	0	1	0	0	1
X	1	2	1	1	5

TABLE XIII

FREQUENCY OF SUPERVISORY RATINGS OF MALE TEACHERS WHO HAD VARIOUS MMPI SCALES CODED AMONG THE TWO LOWEST

Scale Coded Among the Two Lowest	Supervisory Rating				Total
	1. 5	2	2. 5	3	
0	1	5	0	1	7
1	0	5	0	1	6
2	0	1	0	0	1
3	0	1	0	0	1
4	1	3	0	0	4
5	0	1	0	1	2
6	0	0	0	0	0
7	0	4	0	0	4
8	0	5	0	0	5
9	0	2	0	0	2
X	1	2	1	1	5

TABLE XIV

FREQUENCY OF SUPERVISORY RATINGS OF MALE TEACHERS WHO HAD VARIOUS PAIRS OF MMPI SCALES CODED AMONG THE THREE HIGHEST

Pairs of Scales Coded Among the Three Highest	Supervisory Rating				Total
	1.5	2	2.5	3	
01	0	0	0	0	0
02	0	0	0	0	0
03	0	0	0	0	0
04	0	0	0	0	0
05	0	1	0	0	1
06	0	0	0	0	0
07	0	1	0	0	1
08	0	2	0	0	2
09	0	0	0	0	0
0X	0	0	0	0	0
12	0	0	0	0	0
13	0	0	0	0	0
14	0	0	0	0	0
15	0	3	0	0	3
16	0	0	0	0	0
17	0	1	0	0	1
18	0	2	0	0	2
19	0	0	0	0	0
1X	0	0	0	0	0
23	0	2	0	0	2
24	0	0	0	1	1
25	0	2	0	1	3
26	0	2	0	0	2
27	0	0	0	1	1
28	0	0	0	1	1
29	0	0	0	0	0
2X	0	0	0		

TABLE XIV (Continued)

Pairs of Scales Coded Among the Three Highest	Supervisory Rating				Total
	1.5	2	2.5	3	
34	0	1	0	0	1
35	1	2	1	0	4
36	1	1	0	0	2
37	0	0	0	0	0
38	0	0	0	0	0
39	1	1	1	0	0
3X	0	0	0	0	0
45	0	0	0	1	1
46	0	1	0	0	1
47	0	1	0	1	2
48	0	0	0	0	0
49	0	0	0	1	1
4X	0	0	0	0	0
56	1	2	0	0	3
57	1	1	0	0	2
58	1	4	0	0	5
59	1	0	0	0	1
5X	0	0	0	0	0
67	0	1	0	0	1
68	0	1	0	0	1
69	1	0	0	0	1
6X	0	0	0	0	0
78	1	1	0	1	3
79	0	0	0	1	1
7X	0	0	0	0	0
89	0	0	0	0	0
8X	0	0	0	0	0
9X	0	0	0	0	0

depended upon. The addition of one or two more cases could change the picture entirely in many of the categories.

It can be noted that nine of the twenty men had scale 5 (Mf) coded highest of all the scales (see Table X), and thirteen of the twenty had scale 5 coded among the three highest (Table XI). Only two men had a low coding of scale 5 (see Table XIII). Drake[23] says, "A high coding of this scale among men is particularly frequent in an educational setting."

No other scale was coded highest frequently enough to merit special attention. Scale 3 (Hy) was coded among the three highest seven times, all for above average teachers. Scale 7 (Pt) was coded among the three highest six times, two of these being for men who were rated average teachers. Scale 8 (Sc) was among the three highest scales seven times, distributed among the rating categories.

No pair of scales was coded among the highest three scales frequently enough to warrant any generalizations concerning it (see Table XIV). Pairs 15, 25, 39, 56, and 78 appeared three times each. Pair 35 was coded among the three highest four times and 58 was coded among the three highest five times. Four of these times were for teachers with supervisory ratings of 2 and the other was for a teacher with a rating of 1.5.

Five of the twenty men in the group had no scales coded low (more than one standard deviation below the mean). These five appeared in all four of the supervisory rating categories (see Table XII). Three of the group had scale 0 (Si) coded lowest and two each had scales 1 (Hs), 7 (Pt), and 8 (Sc) coded lowest.

Six of the men had scale 0 (Si) coded among the two lowest scales. A low coding of scale 0 for men is "probably indicative of an adequate social adjustment."[24] Five each had scales 1 (Hs) and 8 (Sc) coded among the two lowest. Four men each had scales 4 (Pd) and 7 (Pt) coded among the two lowest.

Because many of the sub-groups in this study (for both men and women) were very small, it was thought that Fisher's exact probability test might be helpful. Siegel[25] states, "The Fisher exact probability test is an extremely useful nonparametric technique for analyzing discrete data (either nominal or ordinal) when the two independent groups are small in size and when the scores from the two groups fall into one or the other of two mutually exclusive classes." The determining of the exact probabilities by means of Fisher's formula is rather laborious, but many of the values may be obtained from a "Table of Critical Values of D in the Fisher-Yates Test."[26] The Fisher test did not add any new information for this study; but it did support some of the previously determined relationships.

Summary

In summary, it can be stated that the analysis of the data for this study indicates the following relationships, —all of which are limited to the population studied or a similar group:

1. A high coding of scale 3 (Hy) for women teachers was positively related to a supervisory rating of outstanding. Although a statement of this relationship does not appear in any of the literature, Professor L. E. Drake, Head of the Student Counseling Center of the University of Wisconsin, has stated orally that in his experience he has found such a relationship to exist.

Thirty-five percent of the small group of men in the study had scale 3 (Hy) coded among the three highest scales. All of these men were rated above average or outstanding by supervisors.

2. Scale 7 (Pt), although coded highest or among the three highest scales for women teachers more frequently than any other scale, did not distinguish to a marked degree between good and poor teachers as indicated by supervisory ratings.

An absence of scale 7 coded high seemed to be somewhat related to high supervisory ratings for women teachers.

Low codings of scale 7 appeared to be positively related to low supervisory ratings for women.

Scale 7 was not coded high for men as frequently, relatively speaking, as for women teachers. It was coded among the lowest two scales more frequently for men than were many of the other scales.

3. For the women teachers in this study, a high coding of scale 5 (Mf) was positively related to supervisory ratings which were above average, whereas a low coding of this scale seemed to be somewhat related to average or below supervisory ratings. There are some indications that more extensive data might reveal more definite relations between high or low codings of scale 5 and supervisory ratings. It can be noted that in the lowest rated group of female teachers, there is no occurrence of scale 5 coded high.

For male teachers, scale 5 (Mf) coded highest appeared in 45 percent of the cases. It was coded among the three highest scales in 65 percent of the cases. Only one man had scale 5 coded among the two lowest scales.

4. A low coding of scale 2 (D) for women teachers was associated with supervisory ratings that were above average.

5. Scale 0 (Si) was coded lowest much more frequently than it was coded highest for women teachers. A low coding seemed to be somewhat related to above average supervisory ratings.

Scales 2 (D) and 5 (Mf) were also coded lowest more frequently than highest for women.

6. There were only four out of 147 female

teachers who had no scales coded low, whereas five of the twenty male teachers had no scale coded low.

It was pointed out in an earlier section that the MMPI was originally designed for and standardized on mental patients, but that it has since been used advantageously on many different more normal groups in a variety of situations.

It should be pointed out that the people in this study, who were given the MMPI, were all University of Wisconsin graduates who were granted a certificate to teach. The University of Wisconsin maintains an extensive testing, counseling, and guidance program and it is more than likely that most of the students with serious personality abnormalities who were not eliminated through the natural processes associated with attending college would be eliminated through the testing and counseling program. At the University of Wisconsin, students who deviate seriously from the normal are sent to the University psychiatrist for counseling or treatment.

Apparently the University of Wisconsin program is effective in eliminating the serious personality cases before they are graduated and sent out to teach. For the 167 cases in this study, there were only two reports from supervisors or principals which were concerned with emotional instability which was serious enough to affect the teacher's work. If the University's program is successful, one would not expect the MMPI to discover many extreme deviations from normality among the group in this study.

The MMPI may be a valuable instrument in helping to eliminate the occasional seriously deviated personality that does graduate and seek a teaching position.

Further experimentation with the MMPI is recommended to see if there isn't some "cut off" point on certain scales, possibly scales 0 (Si), 5 (Mf), and 7 (Pt), or on combinations of scales, a score beyond which would eliminate a student from teacher candidacy. In order to find such points, especially if combinations of scales are to be considered, it would probably be necessary to carry on a study of several years duration in order to get enough individuals in the study.

One serious deficiency or fault is frequently enough to make a person unsuccessful. Success may be due to one or more superiorities, providing a serious deficiency does not offset the superior qualities, or to a number of mediocrities.

The various scales of the MMPI have been referred to frequently by number or by number and abbreviation. It has been pointed out that the scales were first developed for clinical use with hospital patients and they bear titles indicative of deviate behavior. With the increased use of the Inventory for more normal people and further investigation in regard to the interpretation of scores on the various scales, the authors of the Inventory have suggested that the clinical titles and descriptions of the various scales should probably not be used. They suggest calling the scales by number or letter. This has been done throughout this report.

However, for those who are not familiar with the MMPI, brief descriptions of the various scales will be given. This will be done primarily in the language of Drake[27] who describes the scales in terms of the behavior of more normal persons.

Scale 0 (Si) has the title "Social Introversion-Extroversion". This scale was derived and cross-validated by Drake[28,29]. Its title is more or less self explanatory. The scale is intended to measure a characteristic that is thought to be important in college adjustment. Personality patterns with a high coding of scale 0 (Si) have been found among persons showing introvertive characteristics, especially shyness, social insecurity, and social withdrawal. One might expect such a personality characteristic to be detrimental to teaching success. On the other hand, a low coding of scale 0 (Si), which is probably indicative of an adequate social adjustment, may indicate a higher degree of teacher effectiveness. The present study found low codings of scale 0 (Si) for women teachers was associated with above average supervisory ratings.

Scale 1 (Hy) bears the name "Hypochondriasis". Elevation of this scale appeared relatively infrequently with college men and women, probably because persons with problems of a nature which caused a high coding of this scale sought medical aid. A high coding of this scale in a teacher's profile might indicate trouble which needs medical treatment.

The "Depression" scale has the number 2. A high score on scale 2 (D) is indicative of unhappiness and depression. This condition may be associated with problems in the area of human relations, lack of success in a chosen endeavor, or dissatisfaction with some aspect of the immediate environment. Scale 2 coded low with women were related to extroversion or socially outgoing characteristics. The study here reported found low codings of scale 2 were associated with above average supervisory ratings for women.

The operation of scale 3, "Hysteria" (Hy), with college students appeared to be different than with the patients upon whom it was derived. For normal college people, scale 3 coded high appears to be indicative of a lack of social problems. Scale 3 coded high along with certain other scales seems to indicate a variety of tendencies, but in most of them, a good general adjustment seems prevalent. This should contribute to a high degree of teacher effectiveness. The findings of the present study were that such a relationship exists.

The "Psychopathic Deviate" scale, number 4 (Pd), seemed to have little to do with college educated people. Probably persons who are psychopathic deviates get into trouble and are eliminated. Elevation of this scale is associated with delinquency. There is some evidence that scale 4, in combination with other scales, often is associated with antagonism to authority, aggressive behavior, and conflict situations when it is elevated in a man's profile. With women this aggressiveness seems to be expressed in home conflict or in socially outgoing behavior depending upon the rest of the profile.

Scale 5 (Mf) was the least well validated of the basic scales on the MMPI. It bears the title "Masculinity-Femininity" and was designed to assist in the identification of homosexuals. However, the scale failed to distinguish between the homosexual group and normals. A high coding of this scale among men is particularly frequent in an educational setting. It is a complicated scale to interpret, but it is one of the most valuable scales in counseling students. An elevation of scale 5 means different things, depending upon what other scales are elevated with it. A low coding of scale 5 for women was very common. This corresponds to a high coding for men. In many cases a low coding of 5 acted as a suppressor variable for behavior which is suggested by the rest of the profile. In some patterns it appears to intensify or support the rest of the profile. A high coding of scale 5 with women appears to be associated with socially outgoing behavior. This should contribute to teacher effectiveness.

A high coding of scale 6, "Paranoid", seems to occur infrequently with college men except when paired with scale 4, "Psychopathic Deviate". In this case, it differentiated the anxiety sub-group. High coding of scale 6 (Pa) may be an indicator of personal sensitivity and perhaps extreme concern about the reactions of others to something perceived by the counselee as a deficiency. With teachers a high coding of scale 6 may indicate a person who is overly sensitive to the reactions of fellow teachers or even to students.

Scale 7 (Pt) bears the name "Psychasthenia". A high coding of scale 7 appears to be related to general anxiety states. A low coding with men probably indicates good adjustment. With women a high coding of scale 7 also suggests a feeling of insecurity. In general, high codings of this scale suggest personal problems, insecurity, state of anxiety, physical complaints—none of which are conducive to good teaching and personal relationships. The present study found an absence of scale 7 coded high was somewhat related to high supervisory ratings.

Scale 8, "Schizophrenia", when coded high, frequently suggests mental disorganization, vague goals, lacking motivation, general confused state. This appeared to be more serious with women than with men. It would seem that conditions associated with scale 8 (Sc) are a hindrance to effective teaching.

The "Hypomania" scale, number 9 (Ma), when coded high, appeared to indicate aggressiveness or belligerence in men and energetic action of a more socially approved nature in women. The high coding seemed to be modified considerably depending upon which scales were coded with it. For example, a high coding of scale 0 (Si) with a high coding of 9 was associated with social shyness, but when 0 (Si) was low with 9 high, socially outgoing and verbal behavior was suggested. Low codings of scale 9 (Ma) for women were usually indicative of social shyness when either scale 0 (Si) or scale 2 (D) was coded low. In general, the type of personality to be expected in a teacher with scale 9 coded either high or low seems to depend upon what scales are associated with it and in what manner they are related.

It should be emphasized that it is dangerous to associate the clinical names of most of these scales with individuals who may possess a high or low score on them. Perhaps new scales with new names should be developed for use when normal persons are being studied.

SECTION IV

Summary and Conclusions

Statement of the Problem

The purpose of this study was to investigate the relationship, if any, between scores and profile patterns on the MMPI and success in teaching as indicated by supervisory ratings. Personality has been considered by many to be an important factor in a teacher's success. The many aspects of personality are combined in a multitude of ways to make people react differently to various situations. It is thought that some assessment of the different aspects of a teacher's personality may be helpful in predicting that person's success or failure as a teacher. This study was an attempt to discover whether or not there were relationships between aspects of a teacher's personality and his professional success.

Subjects and Instruments

This study was concerned with students who were enrolled as freshmen at the University of Wisconsin during the fall semesters of 1949, 1950, and 1951, who received the University Teacher's Certificate in 1953, 1954, 1955, or 1956, and who subsequently taught in the public schools of Wisconsin. One hundred sixty-seven subjects constituted this population. Only twenty of these persons were males.

Supervisory ratings were obtained for these individuals by submitting a letter and a rating blank to the superintendents, principals, and supervisors of the schools in which they taught.

Measures of the various phases of the teacher's personality were obtained from the results of the group form of the MMPI which had been given to the subjects when they were freshmen. The scores had been coded according to a plan recommended by one of the authors of the Inventory. The MMPI is a widely used instrument that has been found valuable in clinical studies. It is important that its value for this particular purpose should be ascertained.

Method Employed

Although several different statistical techniques were tried with the data of this study, the chi-square test seemed the most appropriate and provided the best analysis. For purposes of investigation the null hypothesis that teachers with different personality profiles would not differ in their supervisory ratings was established. The chi-square technique was used to determine if any particular sub-group differed from the whole population or from certain other sub-groups in respect to supervisory ratings. Since this was not a sampling study, no generalizations to a larger group could be made, but the findings could be useful in predicting the success of future classes of teachers if one were willing to consider the population of this study as a sample of the population of future teachers, or that conditions will remain basically constant.

Comparisons of frequencies, percentages and other aspects of the data which seemed pertinent were also made. Since the number of males in the study was small and it is generally conceded that males differ from females in their personality patterns, the analysis of the data for the male teachers was made separately. This was limited to the type of descriptive analysis mentioned at the beginning of this paragraph.

Findings

The findings of this study briefly summarized were:

1. There were differences in personality patterns among female teachers with different supervisory ratings. A high coding of scale 3 (Hy) for women was positively related to a supervisory rating of outstanding effectiveness. Scale 3 was also coded high very frequently for male teachers, but the number of males was too small to make any conclusions regarding the relationship.

2. Scale 5 (Mf) seems to be positively correlated to supervisory ratings for women teachers. The findings in this respect were not very significant statistically. Scale 5 coded high was the most frequently noted aspect of the male personality profiles. It was coded among the highest three scales 65 percent of the time.

3. There appears to be an inverse relationship between scale 2 (D) and supervisory ratings for women teachers.

4. Scales 0 (Si), 2 (D), and 5 (Mf) were coded lowest more frequently than they were coded highest for the women teachers.

5. Women teachers tended to have scales coded low more frequently than men teachers.

Conclusions

If one wishes to consider the group here studied as representative of future groups of prospective teachers, one might conclude that the MMPI has good possibilities for helping in the prediction of teacher success. While the study has been limited by small numbers in many of the sub-groups, nevertheless, differences in personality profiles did occur for the different supervisory categories.

Although relatively small samples are frequently adequate in statistical studies, in investigations of this nature, because of the large number of sub-groups into which the total group is broken down, the sub-groups are frequently too small to permit of statistical treatment. Such an investigation demands extremely large numbers if a statistical study is to be made. Such studies might be carried on over a period of years and with the cooperation of several large teacher-training institutions.

The natural processes of college life and the efforts of testing, counseling, and guidance personnel tend to eliminate many of the persons with extreme personality deviations from the rolls of candidates for teaching positions. The MMPI can be of help in identifying these individuals. It has potential usefulness in aiding in the prediction of the success of those who do accept teaching positions.

FOOTNOTES

1. Rostker, L. E. "The Measurement of Teaching Ability," Journal of Experimental Education, XIV (1945), p. 50.
2. Lins, L. J. "The Prediction of Teaching Efficiency," Journal of Experimental Education, XV (1946), p. 59.
3. Lins, op. cit.
4. LaDuke, C. V. "The Measurement of Teaching Ability," Journal of Experimental Education, XIV (1945), p. 100.
5. Albert Ellis, Consulting Psychologist, 333 W. 56th Street, New York 19, New York.
6. Buros, Oscar F., (Editor), Fifth Mental Measurements Yearbook (Highland Park, N. J.: The Gryphon Press, 1959), p. 166.
7. Drake, L. E. and Oetting, E. R. An MMPI

Codebook for Counselors (Minneapolis: University of Minnesota Press, 1959), p. 10.

8. Hathaway, J. R., and Monachesi, E. D. *Analyzing and Predicting Juvenile Delinquency with the MMPI* (Minneapolis: University of Minnesota Press, 1953), pp. 13-28.
9. Lins, *op. cit.*
10. Hathaway, S. R. "A Coding System for MMPI Profiles," *Journal of Consulting Psychology*, XI (1947), pp. 334-37.
11. Hathaway, *op. cit.*
12. Du Mas, F. M. "On the Interpretation of Personality Profiles," *Journal of Clinical Psychology* (1947), pp. 57-65.
13. McCall, William A., and Krause, Gertrude R., "Measurement of Teacher Merit for Salary Purposes," *Journal of Educational Research*, LIII (1959), pp. 73-5.
14. Tate, Merle W., and Clelland, Richard C., *Nonparametric and Shortcut Statistics in the Social, Biological, and Medical Sciences* (Danville, Ill.: Interstate Printers and Publishers, 1957), p. 52.
15. Walker, Helen and Lev, Joseph. *Statistical Inference* (New York: Holt, 1953), p. 107.
16. Tate and Clelland, *op. cit.*, p. 71.
17. Cochran, W. G., "The Comparison of Percentages in Matched Samples," *Biometrika*, XXXVII (1950), pp. 256-66.
18. Lindquist, E. F. *Statistical Analysis in Educational Research* (Chicago: Houghton Mifflin Co., 1940), pp. 44-5.
19. An explanation of this method of calculating chi-square may be found in the Appendix.
20. Walker and Lev, *op. cit.*, pp. 105-6.
21. Drake and Oetting, *op. cit.*, pp. 24-5.
22. *Ibid.*, p. 19.
23. *Ibid.*, pp. 22-3.
24. *Ibid.*, p. 15.
25. Siegel, Sidney, *Nonparametric Statistics for the Behavioral Sciences* (New York: McGraw Hill, 1956), pp. 96 ff.
26. Finney, D. J. "The Fisher-Yates Test of Significance in 2 × 2 Contingency Tables," *Biometrika*, XXXV (1948), pp. 149-54.
27. Drake and Oetting, *op. cit.*
28. Drake, L. E., "Scale 0 (Social Introversion)," in *Basic Readings on the MMPI in Psychology and Medicine*, G. S. Welsh and W. G. Dahlstrom, editors (Minneapolis: University of Minnesota Press, 1956).
29. Drake, L. E., and Thiede, W. B. "Further Validation of Scale 0 (Si)," in *Basic Readings on the MMPI in Psychology and Medicine*, G. S. Welsh and W. G. Dahlstrom, editors (Minneapolis: University of Minnesota Press, 1956).

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 4, June 1961)

A COMPARISON OF TEACHERS' RESPONSE PATTERNS ON THE MMPI WITH RESPONSE PATTERNS OF SELECTED NON-TEACHER GROUPS

ELMER CHRISTIAN RIECK*
Wisconsin State College
Eau Claire, Wisconsin

SECTION I

The Problem and Its Importance

Statement of the Problem

THE PURPOSE of this study was to identify and then compare profile patterns of teachers on the group form of the Minnesota Multiphasic Personality Inventory, hereafter referred to as the MMPI, with the profile patterns of selected non-teacher groups of students at the University of Wisconsin. The three main groups studied, divided according to sex, were students who were graduated from the University of Wisconsin with teaching certificates, students with majors in Letters and Science who did not prepare for teaching, and freshmen and sophomores who dropped out prior to starting the junior year. Those who prepared for teaching were further subdivided into academic and special field majors.

Specifically, the researcher will deal with the following four questions:

1. What, if any, MMPI profile patterns are found for students who were graduated from the University of Wisconsin with teaching certificates?
   a. Graduates of the School of Education with academic majors.
   b. Graduates with teaching certificates in special fields—Home Economics, Music, Agriculture, Art Education, and Physical Education.
2. What, if any, MMPI profile patterns are found for students who were graduated from the University of Wisconsin from the College of Letters and Science but did not prepare for teaching?
3. What, if any, MMPI profile patterns are found for students who dropped out of the University prior to starting the junior year?
4. If profile patterns are found, how do they compare among these groups?

It is hoped that the results of the study may prove useful in counseling students in deciding between teaching and other non-teaching careers.

The writer became interested in the MMPI when in student personnel work during the period of 1948-58. During these years he was privileged to see and counsel many students who sought guidance and who requested assistance in personnel, academic, social, and vocational problems. It was his primary function to help facilitate the growth and development of the student of a realistic understanding of himself and his relationship with society.

In the discharge of such functions the use of psychometric data may be very helpful. On the other hand, they may lead, at times, to erroneous conclusions. The fact that psychometric data are often useful in many situations must grow out of their basic validity for the purpose used.

While most of the earlier researches on the MMPI were done in a hospital setting, research has been conducted in a variety of situations with more normal persons. Drake and his associates (10) have made a number of studies with university students.

---

*Only through the cooperation of many people was this study possible. My sincere gratitude and appreciation is expressed to Professor A. S. Barr, the writer's major professor and Chairman of the Reading Committee, and Professors L. E. Drake and H. J. Klausmeier, members of the Reading Committee. They have been extremely generous in giving their time and effort to this investigation. Their continuous guidance, constructive criticism and encouragement is deeply appreciated. My sincere gratitude is also extended to my friends and associates who have encouraged and aided in this study.

Importance of the Problem

At the present time nearly every large university and most of the smaller institutions offer their students some type of guidance service. At the University of Wisconsin, for example, the Student Counseling Center has offered this service to the students since the Center opened in 1944.

One of the important functions of these counseling centers is that of academic and vocational guidance. The work of assisting students in selecting their academic and vocational goals is difficult. Because of the demands on the part of the students for this type of help and the desire on the part of the counselors to be effective, it was necessary that the MMPI be carefully studied so that it may be used as effectively as possible.

The present study was designed to obtain more information about the suitability of using the MMPI with University of Wisconsin students in academic and vocational guidance, especially in the selection of future teachers. The MMPI is a widely used instrument and has been found useful in clinical studies. It is important that its value for this particular purpose should be ascertained.

SECTION II

The Procedures Used in the Investigation

The Design of the Study

The study here reported was of the comparative-descriptive type in which groups of prospective teachers and non-teaching groups were compared upon scales and profile patterns of the MMPI. In an exploratory study such as this, it seemed best to work with a well-defined population about whom precise and complete information could be had. The study concerned students enrolled at the University of Wisconsin during the fall semesters of 1949, 1950, and 1951.

The Minnesota Multiphasic Personality Inventory

The MMPI was chosen as the measuring instrument to be used in this study. It is normally administered to students entering as freshmen during freshmen orientation week at the University of Wisconsin and the profile codes are entered on the cards by the staff of the Student Counseling Center. Profiles and other data were thus obtained for the subjects upon the basis of which comparisons were made of the several profile patterns.

The MMPI is a psychometric instrument developed at the University of Minnesota. It is a product of S. R. Hathaway, Professor and Director of the Division of Clinical Psychology, University of Minnesota, and J. C. McKinley (deceased) formerly Head of the Department of Neuropsychiatry, University of Minnesota. The Inventory was designed to provide, in a single test, scores on all the more important clinical phases of personality. It seemed that a multiphasic personality inventory would be best suited for our purposes.

The instrument consists of 550 different statements covering a wide range of subject matter, from the physical condition of the individual being tested to his morale and attitude (15). In administering the test, the subject is asked to respond to all statements, which are in the first person, as True, False, or Cannot Say. After he has responded to all items in this manner, the responses are scored and tabulated to faciliate this part of the operation. Drake, at the University of Wisconsin Counseling Center, has recorded the significant responses on an IBM machine score sheet. This device saves considerable time and labor in scoring. Experimentation has shown that it is adapted to useful classification and understanding of college students.

Subjects and Data-Gathering Procedures

The students for this study were students at the University of Wisconsin who entered as freshmen in the fall semesters of 1949, 1950, and 1951.

A list of the students who were graduated in 1953, 1954, and 1955 with teaching certificates was obtained from the Office of the Dean of the School of Education. The profiles of all these students who had taken the MMPI in 1949, 1950, and 1951 were used in this study. Of the total of 376 subjects, 99 were males and 277 females. The Korean situation and the fact that a larger percentage of females than males go into the teaching profession accounts for the larger female group.

The total population consisted of 1407 students, 654 males and 753 females. The profile patterns and other data were obtained from the files of the Student Counseling Center.

Graduates from the College of Letters and Science and freshman and sophomore drop-outs were secured from the records on file at the Alumni Records Office. The profiles of all the Letters and Science students who had taken the MMPI in 1949, 1950, and 1951 that were readily available, 278 males and 151 females, compromises the population of this group of approximately 1750. A total of 502 subjects, 277 males and 225 females, were used from the freshman and sophomore drop-out group. This represented all subjects for whom both drop-out records and MMPI's were available.

Inasmuch as this study attempted to identify and then compare profile patterns of students who subsequently become teachers with profile patterns of other selected groups of students, the subjects were divided into four groups as follows:

1. University of Wisconsin graduates with teaching certificates in the academic subject fields.

2. Graduates with teaching certificates in special fields—Home Economics, Music, Agriculture, Art Education, and Physical Education.

3. Students who were graduated from the University of Wisconsin from the College of Letters and Science but did not prepare for teaching.

4. Students who dropped out of the University prior to starting the junior year.

In order to identify common profile patterns among individuals within a group and to compare possible profile patterns among groups, each group must show relatively homogenous behaviors on the MMPI. Therefore, the group must be selected on the basis of as many common underlying traits as possible. In this study the groups were divided in accordance to sex and education and/or vocational goals.

Coding of the Data

The coding system of the profile analysis was that reported by Hathaway (14). The scales were assigned numbers as follows:

0 Si - Social Introversion
1 Ha - Hypochondriasis
2 D - Depression
3 Hy - Hysteria
4 Pd - Psychopathic Deviate
5 Mf - Masculinity-Femininity
6 Pa - Paranoid
7 Pt - Psychasthenia
8 Sc - Schizophrenia
9 Ma - Hyponrenia
X None

All the scales with T scores above 54 were coded high and arranged according to magnitude. Only the three highest scales were used in this study. The scales with T scores of 45 or less were coded low with a maximum, in this study, of two being coded low except in cases where there were ties in which case the first three low scales were included.

Tabulating the Data

The frequency of occurrence of each of the following patterns of scales were tabulated (8):

1. The frequency of each scale coded high among the three highest scales.
2. The frequency of each scale coded low among the two lowest scales. (In case of a tie, the first three low scales were included.)
3. The frequency of each pair of scales coded among the three highest scales.
4. The frequency of each scale coded high of the three highest with each scale coded low of the two lowest.
5. The frequency of each pair of high scales coded among the three highest with each scale coded low of the two lowest.

Comparisons of the profile patterns among the main groups of students were made by Yule's formula for differentiation of groups (25). This formula assumes randomly drawn groups. Strictly speaking, these groups were not randomly drawn. However, for some larger group of Wisconsin graduates, the groups might be regarded as random groups. To make such a statement it must be assumed that conditions remain unchanged in all essential conditions and that these conditions persist for some period of time; also that the types of individuals appearing in the data have an equal opportunity to appear at each successive drawing. The data here are for graduates who entered as freshmen in three separate years, namely, 1949, 1950, and 1951. The statistical test employed in the study answers the question: What are the probabilities that differences as great or greater than those found might have arisen by chance?

SECTION III

Statistical Analysis and Findings

Introduction

The purpose of this study, as stated in Section I, was to identify and compare profile scales and patterns of teachers on the group form of the MMPI with the profile patterns of selected nonteacher groups of students at the University of Wisconsin. The following four questions were asked:

1. What, if any, MMPI profile patterns are found for students who graduated from the University of Wisconsin with teaching certificates?
   a. Graduates of the School of Education with academic majors.
   b. Graduates with teaching certificates in special fields—Economics, Music, Agriculture Art Education, and Physical Education.

2. What, if any, MMPI profile patterns are found for students who were graduated from the col-

lege of Letters and Science but did not prepare for teaching?

3. What, if any, profile patterns are found for students who dropped out of the University prior to starting the junior year?
4. If profile patterns are found, how do they compare among these groups?

In order to answer these questions it is necessary to explain how the data were analyzed and what the statistical findings were.

Yule's Formula for Differentiations of Groups

In Section II, The Design of the Study, the MMPI, the coding of the data, the procedure for tabulating the data, were explained. These data were analyzed through the use of Yule's (25) formula for differentiation of groups. The significant difference between the groups was determined by the use of the formula for the standard error of differences.

$$e_{12}2 = \frac{p_1 q_1}{n^1} + \frac{p_2 q_2}{n^2} \qquad C.R. = \frac{d}{e}$$

where

e_{12} is the standard error of difference

p_1 = the percent of our group having the given code type

q_1 = the percent of the group not having the given code type

p_2 = the percent of the other group having the given code type

q_2 = the percent of that group not having the given code type

n_1 and n_2 = the number of individuals in each group (25)

The size of $e_{12}2$ is an index of the probability of the difference in proportions arising through chance due to the fact that the researcher sampled the total number of persons engaged as explained in Section II. According to Yule, if the difference in proportion does not exceed three times the error of the difference, it may be obliterated by an error of simple sampling on taking fresh samples in the same way from the same material. But for practical purposes the probability of the difference being obliterated in this manner is very little less if the error is twice the difference in proportion.

Six group comparisons were made for each sex, male and female, as follows:

1. Graduates with teaching certificates in the academic subject fields were compared with the graduates with teaching certificates in special fields.

2. Graduates with teaching certificates in the academic subject fields were compared with students who were graduated from the College of Letters and Science but did not prepare for teaching.

3. Graduates with teaching certificates in the academic subject fields were compared with students who dropped out of the University prior to starting the junior year.

4. Graduates with teaching certificates in special fields were compared with students who were graduated from the College of Letters and Science but did not prepare for teaching.

5. Graduates with teaching certificates in special fields were compared with the students who dropped out of the University prior to starting the junior year.

6. Students who were graduated from the College of Letters and Science but did not prepare for teaching were compared with the students who dropped out of the University prior to starting the junior year.

In this study a difference in percent was considered significant by the researcher if it was two or more times as great as its standard error, i.e., it is significant beyond the 5 percent level.

Statistical Findings

The statistical findings are summarized in the tables that follow. They include results for those scales and patterns for which the difference between the groups compared is two or more times as great as the standard error (9). Code types refer to the coding of the scales, either high or low, and combinations of scales, paired high and high with low.

In this investigation, combination of scales paired high with low scales were not tested for significance. The frequencies of occurrence for this code type appear in the Appendix (which will be found in original thesis on file in Library of University of Wisconsin). It was felt that there was an insufficient number of subjects to allow tests for significant differences between these very low frequencies. For the interested researcher some statistical comparisons could be made and other information gained from inspection, if so desired. Drake (9) has reported the frequency of occurrence of a number of patterns found in male clients.

In reporting the findings, the tables show the percentage of the frequency of occurrence of the various code types for all sub-groups compared with the corresponding percentage of the frequency of other sub-groups along with the critical

TABLE I

COMPARISON OF GRADUATES WITH TEACHING CERTIFICATES IN THE ACADEMIC SUBJECT FIELDS WITH THE GRADUATES WITH TEACHING CERTIFICATES IN SPECIAL FIELDS—MALES

Code Types	Percent of Academic Group	Percent of Special Fields Group	Critical Ratio
X High	14. 3*	00. 0	2. 42
7 Low	08. 6*	26. 6*	2. 47
13 Paired High	00. 0	07. 8	2. 33
46 Paired High	00. 0	09. 4*	2. 58
2-9 High with Low	00. 0	06. 3	2. 07
3-2 High with Low	00. 0	07. 8*	2. 33
4-8 High with Low	00. 0	10. 9*	2. 79
6-7 High with Low	00. 0	06. 3	2. 07
Number in each group	35	64	

*These items represent samples of 5 or more cases in one or the other of the compared group N. B. after the number in each case.

TABLE II

COMPARISON OF GRADUATES WITH TEACHING CERTIFICATES IN THE ACADEMIC SUBJECT FIELDS WITH THE GRADUATES WITH TEACHING CERTIFICATES IN SPECIAL FIELDS—FEMALES

Code Types	Percent of Academic Group	Percent of Special Fields Group	Critical Ratio
0 Low	27. 4*	16. 8*	2. 14
5 Low	42. 1*	30. 1*	2. 07
12 Paired High	03. 0*	00. 0	2. 26
0-3 High with Low	03. 0*	10. 6*	2. 38
4-6 High with Low	03. 0*	00. 0	2. 26
5-1 High with Low	02. 4	09. 7*	2. 41
9-5 High with Low	16. 5*	08. 0*	2. 20
Number in each group	164	113	

*These items represent samples of 5 or more cases in one or the other of the compared group N. B. after the number in each case.

TABLE III

COMPARISON OF GRADUATES WITH TEACHING CERTIFICATES IN THE ACADEMIC SUBJECT FIELDS WITH STUDENTS WHO WERE GRADUATED FROM THE COLLEGE OF LETTERS AND SCIENCE BUT DID NOT PREPARE FOR TEACHING—MALES

Code Types	Percent of Academic Group	Percent of Letters and Science Group	Critical Ratio
X High	14. 3*	0. 7	2. 28
8 High	42. 9*	25. 2*	2. 02
0 Low	31. 4*	58. 9*	3. 28
05 Paired High	00. 0	03. 6*	3. 21
13 Paired High	00. 0	07. 2*	4. 65
14 Paired High	00. 0	03. 6*	3. 21
36 Paired High	02. 7	11. 2*	2. 43
46 Paired High	00. 0	03. 6*	3. 21
56 Paired High	02. 9	11. 5*	2. 51
67 Paired High	00. 0	02. 2*	2. 49
0-X High with Low	00. 0	01. 8*	2. 25
1-0 High with Low	00. 0	06. 5*	4. 39
1-7 High with Low	00. 0	02. 5*	2. 66
1-9 High with Low	00. 0	02. 5*	2. 66
2-9 High with Low	00. 0	03. 6*	3. 21
3-1 High with Low	00. 0	05. 8*	4. 14
3-2 High with Low	00. 0	05. 0*	3. 82
3-4 High with Low	00. 0	02. 5*	2. 66
4-0 High with Low	05. 7	16. 9*	2. 47
4-8 High with Low	00. 0	01. 8*	2. 25
4-9 High with Low	00. 0	07. 7*	4. 81
6-0 High with Low	02. 9	10. 1*	2. 14
6-7 High with Low	00. 0	05. 4*	4. 00
7-0 High with Low	00. 0	04. 0*	3. 39
7-1 High with Low	00. 0	02. 5*	2. 66
7-9 High with Low	00. 0	04. 0*	3. 39
8-1 High with Low	00. 0	01. 8*	2. 25
8-6 High with Low	00. 0	03. 6*	3. 21
9-3 High with Low	00. 0	02. 2*	2. 49
9-4 High with Low	00. 0	03. 6*	3. 21
Number in each group	35	278	

---

*These items represent samples of 5 or more cases in one or the other of the compared group N. B. after the number in each case.

TABLE IV

COMPARISON OF GRADUATES WITH TEACHING CERTIFICATES IN THE ACADEMIC SUBJECT FIELDS WITH STUDENTS WHO WERE GRADUATED FROM THE COLLEGE OF LETTERS AND SCIENCE BUT DID NOT PREPARE FOR TEACHING—FEMALES

Code Types	Percent of Academic Group	Percent of Letters and Science Group	Critical Ratio
3 High	28. 0*	39. 1*	2. 09
5 Low	42. 1*	26. 5*	2. 96
78 Paired High	20. 1*	11. 3*	2. 17
1-6 High with Low	02. 4	00. 0	2. 00
2-9 High with Low	04. 9*	00. 0	2. 90
3-X High with Low	02. 4	06. 6*	2. 03
3-1 High with Low	02. 4	07. 9*	2. 20
4-5 High with Low	13. 4*	06. 0*	2. 25
4-9 High with Low	04. 9*	0. 7	2. 32
7-X High with Low	04. 9*	12. 6*	2. 42
7-0 High with Low	06. 7*	02. 0	2. 08
7-5 High with Low	22. 0*	12. 6*	2. 24
8-5 High with Low	15. 9*	07. 9*	2. 22
8-9 High with Low	04. 9*	0. 7	2. 31
Number in each group	164	151	

*These items represent samples of 5 or more cases in one or the other of the compared group N. B. after the number in each case.

TABLE V

COMPARISON OF GRADUATES WITH TEACHING CERTIFICATES IN THE ACADEMIC SUBJECT FIELDS WITH STUDENTS WHO DROPPED OUT OF THE UNIVERSITY PRIOR TO STARTING THE JUNIOR YEAR—MALES

Code Types	Percent of Advanced Group	Percent of Drop-Outs	Critical Ratio
X High	14. 3*	02. 2*	2. 02
7 Low	08. 6	23. 5*	2. 77
02 Paired High	00. 0	02. 9*	2. 87
05 Paired High	00. 0	01. 8*	2. 25
09 Paired High	00. 0	01. 8*	2. 25
13 Paired High	00. 0	04. 7*	3. 69
14 Paired High	00. 0	01. 8*	2. 25
34 Paired High	02. 9	09. 7*	2. 03
37 Paired High	00. 0	02. 2*	2. 49
38 Paired High	00. 0	02. 5*	2. 66
46 Paired High	00. 0	05. 1*	3. 83
49 Paired High	05. 7	15. 2*	2. 24
67 Paired High	00. 0	04. 0*	3. 39
9X Paired High	00. 0	01. 8*	2. 25
0-X High with Low	00. 0	01. 8*	2. 25
0-1 High with Low	00. 0	02. 2*	2. 49
0-2 High with Low	00. 0	02. 2*	2. 49
0-5 High with Low	00. 0	02. 2*	2. 49
0-6 High with Low	00. 0	03. 6*	3. 21
1-2 High with Low	00. 0	02. 5*	2. 66
1-9 High with Low	00. 0	02. 2*	2. 49
2-9 High with Low	00. 0	04. 3*	3. 52
3-2 High with Low	00. 0	04. 3*	3. 52
3-5 High with Low	00. 0	03. 2*	3. 02
4-5 High with Low	00. 0	02. 2*	2. 49
4-8 High with Low	00. 0	05. 1*	3. 83
4-9 High with Low	00. 0	02. 9*	2. 87
5-3 High with Low	00. 0	03. 2*	3. 02
6-7 High with Low	00. 0	05. 1*	3. 83
6-9 High with Low	00. 0	03. 2*	3. 02
7-1 High with Low	00. 0	02. 9*	2. 87
7-5 High with Low	00. 0	02. 5*	2. 66
7-9 High with Low	00. 0	03. 6*	3. 21
8-X High with Low	28. 6*	12. 3*	2. 06
8-5 High with Low	00. 0	02. 2*	2. 49
8-6 High with Low	00. 0	03. 6*	3. 21
9-3 High with Low	00. 0	03. 6*	3. 21
9-4 High with Low	00. 0	01. 8*	2. 25
9-7 High with Low	02. 9	11. 9*	2. 62
Number in each group	35	277	

---

*These items represent samples of 5 or more cases in one or the other of the compared group N. B. after the number in each case.

TABLE VI

COMPARISON OF GRADUATES WITH TEACHING CERTIFICATES IN THE ACADEMIC SUBJECT FIELDS WITH STUDENTS WHO DROPPED OUT OF THE UNIVERSITY PRIOR TO STARTING THE JUNIOR YEAR—FEMALES

Code Types	Percent of Academic Group	Percent of Drop-Outs	Critical Ratio
5 Low	42. 1*	26. 7*	3. 17
7 High	42. 1*	32. 0*	2. 04
9 High	39. 0*	50. 7*	2. 30
29 Paired High	00. 0	02. 2*	2. 24
49 Paired High	07. 3*	14. 2*	2. 23
59 Paired High	04. 9*	11. 1*	2. 30
78 Paired High	20. 1*	12. 4*	2. 00
4-1 High with Low	0. 6	05. 8*	3. 11
6-0 High with Low	06. 1*	12. 9*	2. 34
7-5 High with Low	22. 0*	13. 8*	2. 09
7-9 High with Low	06. 7*	01. 3	2. 57
8-5 High with Low	15. 9*	08. 9*	2. 04
9-1 High with Low	04. 3*	11. 6*	2. 74
9-7 High with Low	05. 5*	11. 6*	2. 19
9-8 High with Low	00. 0	03. 6*	2. 90
Number in each group	164	225	

*These items represent samples of 5 or more cases in one or the other of the compared group N. B. after the number in each case.

TABLE VII

COMPARISON OF GRADUATES WITH TEACHING CERTIFICATES IN SPECIAL FIELDS WITH STUDENTS WHO WERE GRADUATED FROM THE COLLEGE OF LETTERS AND SCIENCE BUT DID NOT PREPARE FOR TEACHING—MALES

Code Types	Percent of Special Fields Group	Percent of Letters and Science Group	Critical Ratio
0 Low	29. 7*	58. 9*	4. 52
8 Low	26. 6*	12. 6*	2. 38
08 Paired High	00. 0	03. 2*	3. 02
14 Paired High	00. 0	03. 6*	3. 21
19 Paired High	01. 6	07. 2*	2. 55
29 Paired High	00. 0	01. 8*	2. 25
39 Paired High	04. 7	13. 7*	2. 69
45 Paired High	04. 7	12. 9*	2. 47
0-X High with Low	00. 0	01. 8*	2. 25
1-0 High with Low	01. 6	06. 5*	2. 28
3-X High with Low	01. 6	05. 8*	2. 00
3-4 High with Low	00. 0	02. 5*	2. 66
3-6 High with Low	00. 0	02. 2*	2. 49
4-0 High with Low	03. 1	16. 9*	4. 42
4-6 High with Low	00. 0	03. 6*	3. 21
4-8 High with Low	10. 9*	01. 8*	2. 29
7-1 High with Low	00. 0	02. 5*	2. 66
8-3 High with Low	00. 0	02. 5*	2. 66
8-9 High with Low	00. 0	02. 5*	2. 66
9-X High with Low	04. 7	11. 2*	2. 00
9-0 High with Low	09. 4*	22. 7*	3. 00
9-3 High with Low	00. 0	02. 2*	2. 49
Number in each group	64	278	

*These items represent samples of 5 or more cases in one or the other of the compared group N. B. after the number in each case.

TABLE VIII

COMPARISON OF GRADUATES WITH TEACHING CERTIFICATES IN SPECIAL FIELDS WITH STUDENTS WHO WERE GRADUATED FROM THE COLLEGE OF LETTERS AND SCIENCE BUT DID NOT PREPARE FOR TEACHING—FEMALES

Code Types	Percent of Special Fields Group	Percent of Letters and Science Group	Critical Ratio
3 High	26.5*	39.1*	2.19
9 High	32.7*	49.7*	2.83
3 Low	13.3*	04.6*	2.40
04 Paired High	03.5	00.0	2.02
39 Paired High	05.3*	13.9*	2.44
5X Paired High	07.1*	01.3	2.24
67 Paired High	11.5*	04.6*	2.00
79 Paired High	10.6*	19.2*	2.20
89 Paired High	00.7	14.6*	3.50
4-6 High with Low	00.0	02.6	2.00
5-1 High with Low	09.7*	03.3*	2.06
5-3 High with Low	04.4*	00.0	2.28
6-0 High with Low	03.5	11.3*	2.52
7-3 High with Low	08.0*	01.3	2.47
7-5 High with Low	23.9*	12.6*	2.33
9-0 High with Low	08.8*	16.6*	2.25
9-7 High with Low	03.5	08.6*	2.17
Number in group	113	151	

*These items represent samples of 5 or more cases in one or the other of the compared group N. B. after the number in each case.

TABLE IX

COMPARISON OF GRADUATES WITH TEACHING CERTIFICATES IN SPECIAL FIELDS WITH THE STUDENTS WHO DROPPED OUT OF THE UNIVERSITY PRIOR TO STARTING THE JUNIOR YEAR—MALES

Code Types		Percent of Special Fields Group	Percent of Drop-Outs	Critical Ratio
X	High	00.0	02.2*	2.49
3	Low	01.6	08.7*	3.07
8	Low	26.6*	13.7*	2.19
02	Paired High	00.0	02.9*	2.87
08	Paired High	00.0	03.6*	3.21
14	Paired High	00.0	01.8*	2.25
29	Paired High	00.0	03.0*	3.36
38	Paired High	00.0	02.5*	2.66
39	Paired High	04.7	11.9*	2.19
5X	Paired High	00.0	01.8*	2.25
58	Paired High	15.6*	04.3*	2.40
0-X	High with Low	00.0	01.8*	2.25
0-1	High with Low	00.0	02.2*	2.49
0-5	High with Low	00.0	02.2*	2.49
1-5	High with Low	00.0	01.8*	2.25
3-6	High with Low	00.0	03.6*	3.21
4-0	High with Low	03.1	11.6*	2.93
4-6	High with Low	00.0	02.9*	2.87
5-3	High with Low	00.0	03.2*	3.02
7-1	High with Low	00.0	02.9*	2.87
8-3	High with Low	00.0	04.0*	3.36
8-5	High with Low	00.0	02.2*	2.49
9-X	High with Low	04.7	11.6*	2.11
9-3	High with Low	00.0	03.6*	3.21
9-6	High with Low	04.7	11.6*	2.11
Number in group		64	277	

---

*These items represent samples of 5 or more cases in one or the other of the compared group N. B. after the number in each case.

TABLE X

COMPARISON OF GRADUATES WITH TEACHING CERTIFICATES IN SPECIAL FIELDS WITH THE STUDENTS WHO DROPPED OUT OF THE UNIVERSITY PRIOR TO STARTING THE JUNIOR YEAR—FEMALES

Code Types		Percent of Special Field Groups	Percent of Drop-Outs	Critical Ratio
9	High	32. 7*	50. 7*	3. 25
0	Low	16. 8*	30. 2	2. 87
9	Low	19. 5*	09. 3*	2. 42
12	Paired High	00. 0	01. 8	2. 02
15	Paired High	00. 0	02. 2*	2. 24
16	Paired High	00. 0	02. 7*	2. 50
23	Paired High	00. 9	04. 9*	2. 35
3X	Paired High	05. 3*	00. 9	2. 00
39	Paired High	05. 3*	11. 6*	2. 10
6X	Paired High	00. 0	02. 7*	2. 03
64	Paired High	02. 7	07. 6*	2. 10
69	Paired High	00. 2	12. 9*	3. 98
0-9	High with Low	10. 0*	03. 1*	2. 40
1-7	High with Low	00. 0	01. 8	2. 02
4-1	High with Low	01. 8	05. 8*	2. 00
4-6	High with Low	00. 0	03. 1*	2. 70
5-9	High with Low	07. 1*	00. 9	2. 50
6-0	High with Low	03. 5	12. 9*	3. 33
6-1	High with Low	01. 8	06. 7*	2. 36
7-5	High with Low	23. 9*	13. 8*	2. 18
9-0	High with Low	08. 8*	20. 0*	2. 90
9-7	High with Low	03. 5	11. 6*	2. 96
Number in group		113	225	

*These items represent samples of 5 or more cases in one or the other of the compared group N. B. after the number in each case.

TABLE XI

COMPARISON OF STUDENTS WHO WERE GRADUATED FROM THE COLLEGE OF LETTERS AND SCIENCE BUT DID NOT PREPARE FOR TEACHING WITH THE STUDENTS WHO DROPPED OUT OF THE UNIVERSITY PRIO TO STARTING THE JUNIOR YEAR—MALES

Code Types	Percent of Letters and Science Group	Percent of Drop-Outs	Critical Ratio
1 High	18. 9*	11. 6*	2. 13
5 High	54. 7*	35. 7*	4. 57
0 Low	58. 9*	23. 8*	8. 97
3 Low	03. 6*	08. 7*	2. 51
5 Low	02. 2*	07. 2*	2. 80
15 Paired High	05. 8*	01. 8*	2. 50
19 Paired High	07. 2*	03. 2*	2. 13
57 Paired High	14. 0*	05. 4*	3. 47
58 Paired High	09. 4*	04. 3*	2. 39
0-5 High with Low	00. 0	02. 2*	2. 49
1-0 High with Low	06. 5*	01. 4	3. 12
1-5 High with Low	00. 0	01. 8*	2. 25
3-1 High with Low	05. 8*	01. 4	2. 80
3-4 High with Low	02. 5*	00. 4	2. 08
3-5 High with Low	00. 4	03. 2*	2. 39
4-8 High with Low	01. 8*	05. 1*	2. 13
4-9 High with Low	07. 7*	02. 9*	2. 54
5-3 High with Low	00. 4	03. 2*	2. 50
5-4 High with Low	04. 0*	01. 1	2. 18
7-0 High with Low	04. 0*	00. 7	2. 58
7-2 High with Low	00. 4	02. 9*	2. 31
7-5 High with Low	00. 0	02. 5*	2. 66
8-2 High with Low	02. 2*	06. 1*	2. 31
9-0 High with Low	22. 7*	15. 5*	2. 11
Number in group	278	277	

*These items represent samples of 5 or more cases in one or the other of the compared group N. B. after the number in each case.

TABLE XII

COMPARIOSN OF STUDENTS WHO WERE GRADUATED FROM THE COLLEGE OF LETTERS AND SCIENCE BUT DID NOT PREPARE FOR TEACHING WITH THE STUDENTS WHO DROPPED OUT OF THE UNIVERSITY PRIOR TO STARTING THE JUNIOR YEAR—FEMALES

Code Types		Percent of Letters and Science Group	Percent of Drop-Outs	Critical Ratio
3	High	39. 1*	27. 1*	2. 42
3	Low	04. 6*	10. 7*	2. 28
04	Paired High	00. 0	02. 7*	2. 50
15	Paired High	00. 0	02. 2*	2. 24
26	Paired High	00. 7	03. 6*	2. 07
2-9	High with Low	00. 0	03. 1*	2. 67
3-2	High with Low	10. 6*	03. 6*	2. 50
5-3	High with Low	00. 0	02. 2*	2. 24
5-9	High with Low	04. 6*	00. 9	2. 03
6-3	High with Low	00. 7	04. 0*	2. 24
7-X	High with Low	12. 6*	05. 8*	2. 18
7-3	High with Low	01. 3	06. 2*	2. 63
7-5	High with Low	02. 6	13. 8*	3. 15
Number in group		151	225	

*These items represent samples of 5 or more cases in one or the other of the compared group N. B. after the number in each case.

ratio of the differences.

The data relative to male graduates with teaching certificates in the academic subject fields as compared with the graduates with teaching certificates in specific fields are given in Table I. Table II makes the same comparisons for females. Tables III and IV show the graduates with teaching certificates in the academic subjects as compared with the students who were graduated from the College of Letters and Science but did not prepare for teaching. Tables V and VI show the data for graduates with teaching certificates in the academic subject fields as compared with the students who dropped out of the University prior to starting the junior year.

Comparisons of graduates with teaching certificates in special fields with students who were graduated from the College of Letters and Science but did not prepare for teaching are shown in Tables VII and VIII. The same comparisons were made of the graduates in special fields with students who dropped out of the University prior to starting the junior year in Tables IX and X.

Tables XI and XII show the data in the same way for the students who were graduated from the College of Letters and Science but did not prepare for teaching as compared with the students who dropped out of the University prior to starting the junior year.

Perhaps the most important and highly significant finding is the fact that statistically significant differences did occur, i. e., personality differences as represented by significant differences in pattern of scales of the MMPI did occur in all groups.

At this stage of the research it is not possible to actually identify the meanings of these personality differences. Further research may identify the meaning of these code types in personality terms. This may point to a better idea of personality variables in the future.

It appears from the data that there were fewer differences among the graduates with teaching certificates in the academic fields as compared with the graduates with teaching certificates in the special fields than between any other groups. The two groups of graduates with teaching certificates seem to be relatively homogeneous. The fact that a grade point average of 2. 5 is required for admittance to the School of Education may possibly account for this finding together with the possibility that the personality characteristics that helped lead to selection of teaching as a career.

The largest number of significant different code type frequencies occurred in the group of students who dropped out of the University prior to starting the junior year. In considering this finding the reader should keep in mind that while a sizeable number of academic females (N = 164) was involved, the number of academic males was only 35. Scale 9 (Ma) coded high and in combination with low coded scales was characteristic of the drop-out group. Many of the significant differences occur in patterns involving scale 7 (Pt), especially when scale 7 is coded high with 5 (Mf) coded low. The male profiles with no scales (X) coded high appears to distinguish the two teaching certificate groups from the drop-outs. More comparisons may be possible through further research.

An examination of the data also shows that more differences in frequencies of code types in male groups occurred than in female groups. Possibly there are several reasons for this. Many counselors have found that college females are not as concerned about their academic careers as are male students. Also, the number of students in the female group was larger than the male group in the groups of graduates with teaching certificates.

SECTION IV

Summary and Conclusions

Statement of the Problem

The purpose of this study was to identify and compare profile patterns of teachers with the profile patterns of selected non-teacher groups of students at the University of Wisconsin on the group form of the MMPI. The three main groups studied, divided according to sex, were students who were graduated from the University of Wisconsin with teaching certificates, students with majors in Letters and Science who did not prepare for teaching, and students who dropped out of the University prior to starting the junior year. Those who prepared for teaching were further subdivided into academic and special field majors.

Conclusions were drawn with reference to four questions:

1. What, if any, MMPI profile patterns are found for students who were graduated from the University of Wisconsin with teaching certificates?
    a. Graduates of the School of Education with academic majors.
    b. Graduates with teaching certificates in special fields—Home Economics, Music, Agriculture, Art Education, and Physical Education.

2. What, if any, profile patterns are found for students who were graduated from the College of Letters and Science but did not prepare for teaching?

3. What, if any, profile patterns are found for

students who dropped out of the University prior to starting the junior year?

4. If profile patterns are found, how do they compare among the groups?

While most of the earlier researches on the MMPI were done in a hospital setting, research has been conducted in a variety of situations with normal persons, especially with university students.

The Design of the Study

The study here reported was of the comparative-descriptive type in which groups of prospective teachers and non-teaching groups were compared upon scales and profile patterns of the MMPI In an exploratory study such as this it seemed best to work with a well defined population about whom precise and rather complete information could be had. The study concerned students enrolled at the University of Wisconsin during the fall semesters of 1949, 1950, and 1951.

The present study was designed to obtain more information about the suitability of using the MMPI with University of Wisconsin students in academic and vocational guidance, especially in the selection of future teachers. The MMPI is a widely used instrument and has been found useful in clinical studies. It is important that its value for this particular purpose should be ascertained.

The data were analyzed through the use of Yule's formula for differentiation of groups.

$$e_{12}2 = \frac{p_1\ q_1}{n^1} + \frac{p_2\ q_2}{n^2}$$

The significance of differences between the groups were determined by the use of this formula.

Difference in profile patterns were found in six comparisons made for each sex.

Findings

The detailed findings briefly summarized were:

1. The most important and highly significant finding is the fact that statistically significant differences did occur among all groups.

2. It appears from the data that there were fewer differences among the graduates with teaching certificates in the academic fields as compared with the graduates with teaching certificates in the special fields than between any other groups.

3. The largest number of significant code type frequencies occurred in the group of students who dropped out of the University.

4. More differences in frequencies of code types in male groups occur than in female groups.

Conclusions

This study seems to indicate that significant differences in profile patterns on a personality test can be found between groups of college students with different academic histories. While the study has been limited by the small numbers of persons in some sub-groups, nevertheless, personality differences as represented by significant differences in patterns of scales of the MMPI did occur.

Research is being continued to determine what the profile patterns may mean. In this respect, counseling experience and further research may help.

There are many interesting and difficult problems in research design for this area. Group research techniques may be employed in building a usable formula or to make useful generalizations. Our ultimate concern, however, is with individuals.

Further research is needed in this area to determine what the profile patterns may mean and to find a measure that predicts early and adequately the personality variables involved in later success.

BIBLIOGRAPHY

1. Barr, A. S. "Problems Associated with the Measurement and Prediction of Teacher Success," Journal of Educational Research, LI (May 1958), pp. 695-99.
2. Barr, A. S. "The Measurement and Prediction of Teacher Efficiency: A Summary of Investigations," Journal of Experimental Education, XVI (June 1948), pp. 203-83.
3. Barr, A. S., et al. "The Measurement and Prediction of Teaching Efficiency," Review of Educational Research, XXV (1955), pp. 261-69.
4. Barr, A. S., and Lois Douglas. "The Pre-Training Selection of Teachers," Journal of Educational Research, XXVIII (October 1934), pp. 100-6.
5. Blum, L. P. "A Comparative Study of Students Preparing for Five Selected Professions Including Teaching," Teacher Personnel Research, Monograph No. 3 (Madison, Wis.: Dembar Publications, 1947), pp. 31-65.
6. Carlile, A. B. "Predicting Performance in the Teaching Profession," Journal of Educational Research, XLVII (May 1954), pp. 641-68.
7. Cronbach, Lee J. Essentials of Psychological Testing (New York: Harper and Brothers

1949), pp. 305-67.

8. Drake, L. E. "MMPI Profiles and Interview Behavior," Journal of Counseling Psychology, III (1954), pp. 92-5.
9. Drake, L. E. "Interpretation of MMPI Profiles in Counseling Male Clients," Journal of Counselling Psychology, III (1956), pp. 83-8.
10. Drake, L. E., and Oetting, E. R. "An MMPI Pattern and a Suppressor Variable Predictive of Academic Achievement," Journal of Counseling Psychology, IV (1957), pp. 245-47.
11. Dressel, P. L., and Mann, W. "Appraisal of the Individual," Review of Educational Research, XXI (1951), pp. 115-31.
12. Good, Carter V., Barr, A. S., and Scates, D. E. The Methodology of Educational Research (New York: D. Appleton-Century Co., 1938), pp. 648-78.
13. Harmon, L. R., and Wiener, D. N. "Use of the Minnesota Multiphasic Personality Inventory in Vocational Advisement," Journal of Applied Psychology, XXIX (1945), pp. 131-41.
14. Hathaway, S. R. "A Coding System for MMPI Profiles," Journal of Counseling Psychology, XI (1947), pp. 334-37.
15. Hathaway, S. R., and Monachesi, E. D. Analyzing and Predicting Juvenile Delinquency with the MMPI (Minneapolis: University of Minnesota Press, 1953), pp. 13-23.
16. Hawkes, G. R. "Use of the Minnesota Multiphasic Personality Inventory in Screening College Students for Counseling Purposes," Journal of Educational Psychology, XLI (February 1950), pp. 116-21.
17. Klausmeier, H. J. Teaching in the Secondary School (New York: Harper and Brothers, 1958).
18. Klausmeier, H. J. Desirable Characteristics of the High School Teacher of Psychology, paper presented to Symposium, American Psychological Association, 1958.
19. Klausmeier, H. J., and Swanson, D. J. "The Development of a Functional Course in Educational Psychology for Teachers," Journal of Educational Psychology, XLI (1950), pp. 449-72.
20. McDaniel, Henry B. Guidance in the Modern School (New York: Dryden Press, 1950).
21. Meehl, P. E. Clinical versus Statistical Prediction (Minneapolis: University of Minnesota Press, 1954), pp. 17-22.
22. Michaelis, John U. "The Prediction of Success in Student Teaching from Personality and Attitude Inventories," University of California Publications in Education, XI (1954), pp. 415-81.
23. Ringness, T. A. "Relationship Between Certain Attitudes Towards Teaching and Teaching Success," Journal of Experimental Education, XXI (September 1952), pp. 1-55.
24. Welsh, G. S., and Dahlstrom, W. G., Editors, Basic Readings on the MMPI in Psychology and Medicine (Minneapolis: University of Minnesota Press, 1956).
25. Yule, G. U. An Introduction to the Theory of Statistics, 8th edition, revised (London: Griffin and Co., 1927), pp. 254-72.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 4, June 1961)

A COMPARISON OF THE READABILITY OF VOCATIONAL AGRICULTURE REFERENCE BOOKS WITH THE READING ABILITY OF THE STUDENTS USING THEM*

RALPH R. BENTLEY and R. EDWARD GALLOWAY**
Purdue University

Introduction

TEACHERS OF vocational agriculture often complain that their students do not understand what they read. In vocational agriculture, as in other high-school subjects, reading is one of the most effective and efficient tools of learning. However, the usefulness of this tool is lost if the material to be read does not reasonably match the capability of the reader. If the material is too difficult, the student will not comprehend; if it is too simple, the student may be insulted and/or bored. In either case, learning suffers.

Few studies have been reported which attempted to match the readability of text or reference materials with the reading ability of high-school students. Although the readability of a large number of agricultural reference materials has been reported, no extensive study was found which compared their readability with the tested reading ability of the students actually using them.

Purpose

It was the purpose of this study to compare the readability of widely used agricultural reference books with the reading ability of the vocational agriculture students using them. In order to make this comparison, it was necessary to measure both the readability of vocational agriculture reference books and the reading ability of students of vocational agriculture who were using them.

Procedure

Selection of Cooperating Schools. One relatively large high school with an enrollment of over 100 students and one relatively small high school with an enrollment of under 100 students were chosen in each of the twelve vocational agriculture districts in Indiana. Large and small schools were selected in each district on the basis of three criteria: 1) the school had a department of vocational agriculture approved by the Indiana State Department of Public Instruction, 2) the teacher of vocational agriculture had taught at least one year in the school, and 3) both the teacher of vocational agriculture and his administrator expressed their willingness to cooperate in the study. The basic data for this study were obtained by testing the reading ability of the vocational agriculture students in these schools, and by measuring the readability of the agricultural reference books in their libraries.

Measurement of Readability. Responses of the cooperating teachers of vocational agriculture to a survey indicated that almost 100 reference books were available in their libraries in quantities large enough for class use. Ten of the books found most frequently in the cooperating schools were chosen for readability measurement. These books were A) Your Farming Program, B) Shopwork on the Farm, C) Profitable Farm Management, D) Beef Production, E) Livestock and Poultry Production, F) Approved Practices in Beef Production, G) Veterinary Guide for Farmers, H) Crop Production, I) Productive Soils, and J) Feeds and Feeding (abr.).

The Dale-Chall Readability Formula was used to measure the readability or reading difficulty of these ten books. This meaurement yielded a raw score for each of twenty or more sample portions of reading material from each book. The readability of each book was expressed as a mean and a standard deviation of the sample scores. Analysis of variance was used to test differences among the ten books.

---

* Adapted from a Ph. D. dissertation completed at Purdue University.

**The authors wish to express their gratitude to the Purdue Research Foundation for its support of this study.

Measurement of Reading Ability. Two tests were used to measure the reading ability of vocational agriculture students. The Cooperative English Tests, Test C1, "Reading Comprehension", was used to measure general reading ability, and the Iowa Tests of Educational Development, Test 6, "Ability to Interpret Reading Materials in the Natural Sciences", was chosen as the most appropriate test of ability to read technical materials in agriculture. The Verbal Battery and the Non-verbal Battery of the Lorge-Thorndike Intelligence Tests series were used to determine the mental ability of these students. These tests were administered by the teacher of vocational agriculture in each high school during his regular vocational agriculture class periods. After the scores for each grade from each of the tests were analyzed by analysis of variance, the means and standard deviations were calculated for each school and/or groups of non-differing schools, depending upon the results of the significance tests. These means and standard deviations were converted to grade level units.

Results

The results of this study, designed to compare the readability of certain agricultural reference books with the reading ability of the students using them, are presented under three headings: 1) readability of agricultural reference books, 2) reading ability of vocational agriculture students, and 3) comparison of book readability with student reading ability. In each of these three sections, hypotheses are stated, an analysis of relevant data is presented, and conclusions are drawn. The .05 level of significance was used.

Readability of Vocational Agriculture Reference Books. The raw Dale-Chall scores for sample portions of reading material in each of the ten selected books furnished the basic readability data. Bartlett's test was used to test the homogeneity of within-book variances and a one-way analysis of variance was used to test the significance of differences among the mean Dale-Chall readability scores for the ten books. These tests indicated that significant differences existed both among the variances and the means for the ten books. Therefore, each book was treated individually in the comparison with student reading ability.

Both the mean and the standard deviation readability for each book was converted from raw Dale-Chall score units to grade-level units. This made a direct comparison of readability and reading ability scores possible.

Reading Ability of Vocational Agriculture Students. The raw scores for both the Cooperative Reading Test and the Iowa Reading Test furnished basic reading ability data. The data for each of these tests were analyzed by two-way analysis of variance and presented separately for each of the four high-school grades.

It was hypothesized that the reading ability of ninth-grade students did not differ significantly with respect to the size and location of the schools attended. The results of the analysis of variance for both of the reading tests and the significance of the obtained F values are shown in Table I.

The reading ability of ninth-grade students as measured by the Cooperative and Iowa Reading Tests (Table I) did not differ significantly among the cooperating schools when they were grouped by district and size. Thus it was concluded that the differences among districts and between sizes could have occurred by chance and that these students belong to the same population with respect to reading ability as measured by these two tests. Population means and standard deviations for the Cooperative and the Iowa Reading Tests are shown in Table II.

It was hypothesized that the reading ability of tenth-grade students did not differ significantly with respect to the size and location of the schools attended. The results of the analysis of variance for each test and the significance of the obtained F values are shown in Table III.

The reading ability of tenth-grade students as measured by the Cooperative Reading Test did not differ significantly among the cooperating schools when they were grouped by district and size (Table III). Thus it was concluded that the differences among districts and between sizes could have occurred by chance and that these tenth-grade students of vocational agriculture belong to the same population with respect to their reading ability as measured by the Cooperative test. Table III also shows that ninth-grade students are significantly different from district to district on reading ability as measured by the Iowa test.

The significant F values for interaction in Table III indicate that unique effects of school size occur in some districts for reading ability as measured by the Iowa test. Therefore, it was concluded that students in the various districts belong to different populations and that students in individual schools belong to different populations with respect to their reading ability as measured by the Iowa test. The population mean and the standard deviation for the Cooperative Reading Test are shown in Table IV. The means and standard deviations for the Iowa Reading Test are given in Table V for students in each large and small school by districts.

It was hypothesized that the reading ability of eleventh-grade students of vocational agriculture did not differ significantly with respect to the size and location of the schools they attended. The results of the analysis of variance for both reading tests and the significance of the obtained F values are shown in Table VI.

TABLE I

TWO-WAY ANALYSIS OF VARIANCE OF READING SCORES MADE BY NINTH-GRADE STUDENTS OF VOCATIONAL AGRICULTURE

Source of Variance	Weighted Sums of Squares	d. f.	Mean Square	Significance of F Value*
		Cooperative Reading Test		
District	1640. 4132	10	164. 0413	N. S.
Size	21. 3928	1	21. 3928	N. S.
Interaction	1168. 760	10	116. 8760	N. S.
Within	20782. 239	194	107. 1249	
		Iowa Reading Test		
District	389. 4164	10	38. 9416	N. S.
Size	58. 1405	1	58. 1405	N. S.
Interaction	258. 219	10	25. 822	N. S.
Within	4425. 575	191	23. 1705	

*N. S. indicates that the observed F was not significant at the . 05 level.

TABLE II

MEANS AND STANDARD DEVIATIONS FOR THE COOPERATIVE AND IOWA READING TESTS FOR NINTH-GRADE STUDENTS IN TWENTY-TWO SCHOOLS

Tests	Mean Grade Level of Ability	Grade Levels of S. D.
Cooperative Reading	7. 5	2. 5
Iowa Reading	8. 7	2. 0

TABLE III

TWO-WAY ANALYSIS OF VARIANCE OF READING SCORES MADE BY TENTH-GRADE STUDENTS OF VOCATIONAL AGRICULTURE

Source of Variance	Weighted Sums of Squares	d. f.	Mean Square	Significance of F Value*
		Cooperative Reading Test		
District	2561.5332	10	256.1533	N. S.
Size	191.1195	1	191.1195	N. S.
Interaction	2274.980	10	227.498	N. S.
Within	24476.388	176	139.0704	
		Iowa Reading Test		
District	482.5512	10	48.2551	$P < .05$
Size	.010043	1	.010043	N. S.
Interaction	641.842	10	64.1842	$P < .01$
Within	4134.81	174	23.7633	

*N. S. indicates that the observed F was not significant at the .05 level.

TABLE IV

MEAN AND STANDARD DEVIATION FOR THE COOPERATIVE READING TEST FOR TENTH-GRADE STUDENTS IN TWENTY-TWO SCHOOLS

Test	Mean Grade Level	Grade Levels of Standard Deviation
Cooperative Reading	8.3	3.3

TABLE V

MEANS AND STANDARD DEVIATIONS FOR THE IOWA READING TEST FOR TENTH-GRADE STUDENTS BY DISTRICT AND SCHOOL SIZE

District	Large		Small	
	Mean Grade Level	Standard Deviation	Mean Grade Level	Standard Deviation
	Iowa Reading			
I	10.0	2.8	7.6	.9
II	9.4	1.2	11.1	4.2
III	9.8	2.3	11.9	4.6
IV	11.1	4.9	10.4	2.8
V	8.6	1.0	10.8	3.2
VI	11.1	3.8	12.4	5.3
VII	11.1	3.0	8.0	0.3
VIII	9.8	2.3	9.0	2.3
IX	8.6	1.8	9.7	1.6
XI	9.3	3.1	12.2	3.7
XII	9.6	2.9	8.5	2.7

The reading ability of eleventh-grade students as measured by the Cooperative and Iowa tests did not differ significantly among the twenty-two cooperating schools when they were grouped by district and size. Thus it was concluded that the differences observed among districts and between sizes could have occurred by chance and that these students belong to the same population with respect to their abilities as measured by these two tests.

Population means and standard deviations for the Cooperative and the Iowa Reading Tests are shown in Table VII.

It was hypothesized that the reading ability of twelfth-grade students did not differ significantly with respect to the size and location of the schools they attended. The results of the analysis of variance for each of the reading tests and the significance of the obtained F values are shown in Table VIII.

The reading ability of twelfth-grade students as measured by the Cooperative and the Iowa Reading Tests did not differ significantly among districts but did differ significantly between large and small schools. Therefore, it was concluded that the differences between large and small schools probably did not occur by chance and that the twelfth-grade students in large schools belong to a different population with respect to the Cooperative and the Iowa Reading Tests. Means and standard deviations are shown in Table IX for twelfth-grade students on the Cooperative and Iowa Reading Tests by large and small schools.

Comparison of Book Readability with Student Reading Ability. In order to determine how appropriate vocational agriculture reference books are for the students who use them, comparisons were made on the basis of the readability grade levels of reference books as determined by the Dale-Chall formula and the grade levels of reading ability of students as measured by the Cooperative and Iowa Reading Tests. Both mean and standard deviation grade levels were used in these comparisons. Figure 1 shows these means and standard deviations graphically.

The upper portion of Figure 1 presents the data regarding ten agricultural reference books. Each bar in this portion of the figure shows the mean and the range, minus one to plus one standard deviation, in grade levels of book readability. For example, the first bar in Figure 1 shows that the book, Your Farming Program, has eighth-grade mean readability and that two-thirds of the reading material in this book falls between 6. 4 and 9. 6 grade levels.

The lower portion of Figure 1 graphically presents information that was shown in Tables II, IV, V, VII, and IX about the reading ability of students. Each bar in this portion of the figure shows the mean and the range in grade levels of student reading ability from minus one to plus one standard deviation. Thus, the ninth-grade students of vocational agriculture in the twenty-two cooperating schools had mean general reading ability of grade 7. 5 as measured by the Cooperative Reading Test and two-thirds of these students had a reading ability range of grade 5. 0 and 10. 0.

It will be noted that the mean reading ability of students for each grade level was higher when measured by the Iowa Reading Test than when measured by the Cooperative Reading Test. It will also be noted that mean student reading ability was higher for each succeeding high-school grade except for twelfth-grade students in the large high schools. However, reading ability, as measured by these tests, fails to increase a full grade level for students in each succeeding high-school grade.

Since there was significant interaction among schools with respect to the Iowa Reading Test scores for tenth-grade students, it was not permissable to pool these scores for all tenth-grade students. To show the extreme range of reading ability of tenth-grade students, the highest and the lowest means and their respective standard deviations are shown in Figure 1.

In general, Figure 1 shows that the means for book readability are higher than the means for student reading ability. The grade level means for the ten books studied range from grades eight to eleven with standard deviations ranging from 1. 6 to 2. 7 grade levels. The student grade level means for scores on the Cooperative Reading Test range from 7. 5 to 9. 9 with standard deviations ranging from 2. 5 to 3. 3.

In the following paragraphs the readability of the books as measured by the Dale-Chall formula will be compared with the reading ability of vocational agriculture students as measured by the Cooperative and the Iowa Reading Tests. Both readability and reading ability will be expressed as grade levels. For example, a pupil with ninth-grade reading ability as measured by the Cooperative Reading Test has a score equal to the median score made on this test by students in the publisher's norm group who were halfway through their ninth year of school. Ninth-grade readability implies that the reading material can be read by students who are halfway through the ninth grade. These comparisons are presented separately for each of the four high-school grades.

Figure 1 shows that all ten of the books studied have mean readability scores above the mean reading ability for ninth-grade students as measured by the Cooperative Reading Test. Although these same students scored a grade higher on the Iowa Reading Test, nine of the ten books are still, on the average, too difficult for these ninth-grade students.

The overlapping of standard deviation bars in in Figure 1 shows that while the most able ninth-

TABLE VI

TWO-WAY ANALYSIS OF VARIANCE OF READING SCORES MADE BY ELEVENTH-GRADE STUDENTS OF VOCATIONAL AGRICULTURE

Source of Variance	Weighted Sums of Squares	d. f.	Mean Square	Significance of F Value*
		Cooperative Reading Test		
District	1366. 0296	10	136. 6030	N. S.
Size	254. 10000	1	254. 10000	N. S.
Interaction	1946. 477	10	194. 6477	N. S.
Within	16080. 218	156	103. 0783	
		Iowa Reading Test		
District	69. 6304	10	6. 9630	N. S.
Size	47. 3594	1	47. 3594	N. S.
Interaction	388. 373	10	38. 8373	N. S.
Within	3777. 077	158	23. 9056	

*N. S. indicates that the observed F was not significant at the . 05 level.

TABLE VII

MEANS AND STANDARD DEVIATIONS FOR THE COOPERATIVE AND IOWA READING TESTS FOR ELEVENTH-GRADE STUDENTS IN TWENTY-TWO SCHOOLS

Tests	Mean Grade Level	Grade Levels of Standard Deviation
Cooperative Reading	9. 1	2. 6
Iowa Reading	10. 0	3. 0

TABLE VIII

TWO-WAY ANALYSIS OF VARIANCE OF READING SCORES MADE BY TWELFTH-GRADE STUDENTS OF VOCATIONAL AGRICULTURE

Source of Variance	Weighted Sums of Squares	d. f.	Mean Square	Significance of F Value*
		Cooperative Reading Test		
District	1127.5896	10	112.2590	N. S.
Size	880.5533	1	880.5533	P <. 01
Interaction	1015.1126	10	101.5113	N. S.
Within	14384.024	132	108.9699	
		Iowa Reading Test		
District	423.3408	10	42.3341	N. S.
Size	138.4966	1	138.4966	P <. 05
Interaction	232.851	10	23.2851	N. S.
Within	4422.243	132	33.5018	

*N. S. indicates that the observed F was not significant at the . 05 level.

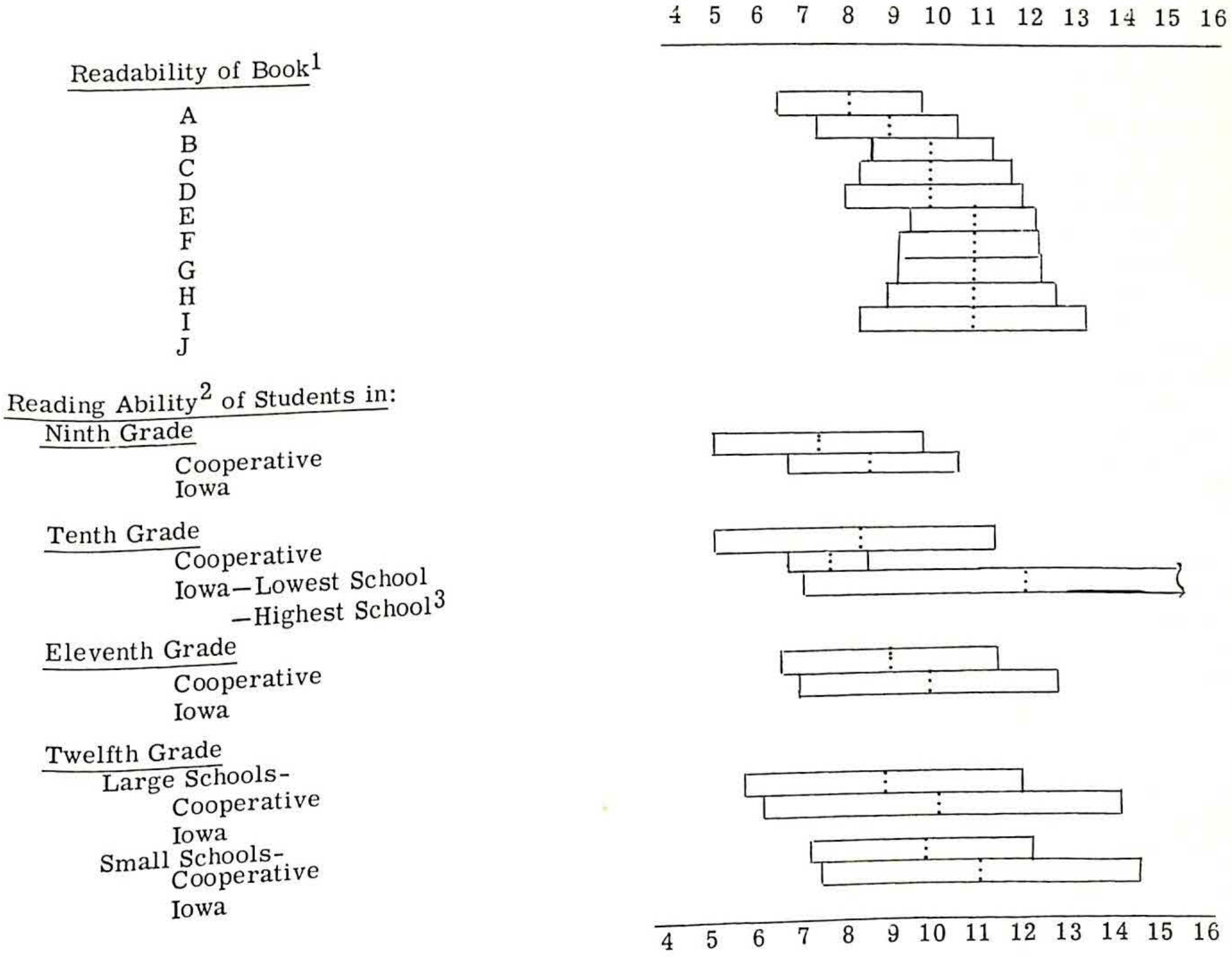
TABLE IX

MEANS AND STANDARD DEVIATIONS FOR THE COOPERATIVE AND IOWA READING TESTS FOR TWELFTH-GRADE STUDENTS SHOWN BY LARGE AND SMALL SCHOOLS

Tests	Large		Small	
	Mean Grade Level	Grade Level of S. D.	Mean Grade Level	Grade Level of S. D.
Cooperative Reading	9. 0	3. 3	9. 9	2. 7
Iowa Reading	10. 3	4. 2	11. 2	3. 7

FIGURE I

GRAPHIC ILLUSTRATION OF MEANS AND STANDARD DEVIATIONS FOR DALE-CHALL READABILITY OF TEN AGRICULTURAL REFERENCE BOOKS AND FOR READING ABILITY OF VOCATIONAL AGRICULTURE STUDENTS AS MEASURED BY THE COOPERATIVE AND IOWA READING TESTS



---

[1] Titles of these ten books are found in the list of books at the end of this study.

[2] Reading ability as measured by the Cooperative and the Iowa reading tests previously described.

[3] Since no pooling was permissable here, only the extreme schools are shown.

grade readers can be expected to read all but the most difficult portions of these ten books, the least able readers could be expected to read only the easiest portions of books A and B.

It is shown in Figure 1 that nine of the ten books have mean readability scores higher than the mean reading ability scores for the tenth-grade students as measured by the Cooperative Reading Test. When compared to mean reading ability as measured by the Iowa Reading Test, the average reading ability for each of the ten books exceeds the lowest tenth-grade class mean and is exceeded by the highest tenth-grade class mean.

The standard deviation bars show that the most able tenth-grade students can be expected to read a major part of the material in each of these ten books and that all but the poorest readers can be expected to read selected portions of these books.

Figure 1 shows that the mean readability of two of the ten books was below the average reading ability of eleventh-grade students and of twelfth-grade students in large schools as measured by the Cooperative Reading Test.

Eight books were above their reading ability. When reading ability was measured by the Iowa test, only half of the books were too difficult, and two were somewhat too easy for these students.

The standard deviation bars indicate that the more able readers in the eleventh grade and in the twelfth grade of large schools can be expected to read all except the most difficult portions of these ten books and that these students will possibly find much material in books A through E too easy. The poorest eleventh-grade readers can be expected to read most of books A and B and the easier portions of the other eight books, while the least able twelfth-grade readers in large schools will probably experience difficulty in reading any of these books.

Figure 1 shows that the twelfth-grade students in the small schools demonstrate reading ability about one grade level higher than the twelfth-grade students in large schools. When compared on their Cooperative Reading Scores, the ability of these students was above the mean readability of books A and B; was about equal to the readability of books C, D, and E; and was one grade level below the means for books F through J. On the basis of their Iowa scores, these students had average ability slightly above the mean readability for each of the ten books.

When the standard deviation bars for the books were compared with those for the twelfth-grade students in small schools, Figure 1 shows that the most capable readers in this group can be expected to read practically all of the material in the ten books with ease. However, these books contain few portions which the least able readers can read.

Conclusions and Implications

The following conclusions were drawn on the basis of reading ability data obtained from the vocational agriculture students in cooperating schools and readability data from reference books selected from the libraries of those schools. In the interpretation of these conclusions, one should be mindful of the limitations imposed by the design and measurements used in this study.

1. Each of the vocational agriculture reference books had a mean readability appropriate for students of average reading ability in one of the four high-school grades.
2. Twelfth-grade students, especially those in large schools, were characterized by their low reading ability.
3. In general, reading ability did not differ with respect to school size and location.
4. In general, verbal mental ability did not differ with respect to school size and location.
5. Nonverbal mental ability varied widely among individual schools.
6. Reading ability varied widely both among vocational agriculture classes and within classes at a given grade level.
7. Students of vocational agriculture scored from one to two grade levels higher on the Iowa Test of Reading in the Natural Sciences than they did on the Cooperative General Reading Test.
8. Vocational agriculture students had mean reading abilities which ranged from zero to three grade levels below their peers on the basis of publisher's norm groups.
9. Although there was wide variation from school to school in the mental ability of vocational agriculture students, in general, their mental ability was similar to that of their peers as represented by publisher's norm groups.
10. In general, the agricultural reference books used by the students of vocational agriculture tended to be too difficult for their reading ability.

The findings and conclusions of this study appear to have several implications for workers in the field of education. These are listed as follows:

1. It seems apparent that if a single vocational agriculture textbook were selected for use by students in a given grade, some students would experience difficulty reading any but the most readable portions and some would find all but the most difficult portions distressingly easy. Therefore, it seems desirable to use a variety of reference materials with varied readability in each agricultural subject area.
2. Teachers cannot assume that a student is an average reader because he possesses average

mental ability.

3. When authors and publishers prepare agricultural reference books, they should give special attention to book readability and the reading ability of vocational agriculture students.

4. The fact that vocational agriculture students consistently scored higher on the test of ability to read natural science materials than on the general reading tests suggests that the ability to read agricultural references may be specific and different from general reading ability.

5. Teachers of vocational agriculture should include readability as a factor in the selection of reference books.

6. Teachers of vocational agriculture should give special attention to the reading problems encountered by their students.

TITLES OF AGRICULTURE REFERENCE BOOKS USED IN THIS STUDY

A. Carsie Hammonds, and W. R. Tabb, Your Farming Program (Lexington, Ky.: Hurst Printing Co., 1951).

B. Mark M. Jones, Shop Work on the Farm (New York: McGraw-Hill Book Co., 1945).

C. James E. Hamilton and W. R. Bryant, Profitable Farm Management (New York: Prentice Hall, 1956).

D. Ronald V. Diggins and Clarence E. Bundy, Beef Production (New York: Prentice Hall, 1953).

E. Clarence E. Bundy, Livestock and Poultry Production (New York: Prentice Hall, 1953).

F. Elwood M. Juergenson, Approved Practices in Beef Production (Danville, Ill.: The Interstate, 1958).

G. G. W. Stamm, Veterinary Guide for Farmers (Chicago: Windsor Press, 1950).

H. Henry L. Algren and Richard J. Delorit, Crop Production (New York: Prentice Hall, 1953).

I. Wilbert Walter Weir; Productive Soils (Chicago: J. B. Lippincott Co., 1946).

J. Frank B. Morrison, Feeds and Feeding, abridged (Ithaca, New York: The Morrison Publishing Co., 1958).

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 4, June 1961)

THE RELATIONSHIP OF CERTAIN LETTER FORM VARIANTS IN THE HANDWRITING OF FEMALE SUBJECTS TO THEIR EDUCATION, I.Q. AND AGE

LAWRENCE EPSTEIN, HUNTINGTON HARTFORD
IRVING TUMARKIN*
Columbia University
College of Physicians and Surgeons
Handwriting Institute, Inc. **

THE MODEL of good penmanship has remained the same throughout the United States for many years, despite the fact that the methods used to teach handwriting have undergone considerable change. As a result, this basic handwriting model, which has been called the "school-copy" model by Lewinson and Zubin (6), may be used as a yardstick to evaluate the tendency of individuals to deviate from what they first learned.

Everyone who writes has deviated somewhat from the "school-copy" form by altering the forms and size ratios of his letters in an idiosyncratic way. It is for this reason that signatures can be accepted as legal identification all over the world. Furthermore, it is the fact that each individual learns to write in a unique but consistent way which underlies the science of graphology.

Two essential reasons that each individual departs from the "school-copy" form are to achieve efficiency and speed. Naturally, the demand for legibility remains even after the demand for rigid reconstruction of the school-copy form is relaxed. Typically, handwriting is not taught to students over 14 years of age and thus, it is then that the tendency to vary accelerates. Many students crystallize their own writing styles in high school or, perhaps, in college when they are called upon to take lecture notes at high speeds.

If handwriting typically departs from the school-copy model, what then is reflected in those adult handwritings which closely adhere to, rather than depart from, the school-copy model? If we regard such samples as cases of arrested handwriting development it would seem to follow that such handwritings may also reflect an arrested intellectual and/or emotional development.

In a recent study, Epstein and Hartford (4) found that in a group of male medical students the retention of the superfluous initial upstrokes on the letters u, v, w, and y[1] was, in fact, related to emotional immaturity. In another recent unpublished study, Linton, Epstein and Hartford found the use of these initial upstrokes to be associated "with relative inadequacy and lack of energy in intellectual functioning and other forms of coping with the environment. . ."

In our survey of earlier graphological studies, we can find but few studies dealing with the relationship between handwriting and intelligence in adult subjects.

Thorndike (11) and Omwake (7) have reported virtually zero correlations between scales of handwriting quality[2] and adult intelligence. These results, in effect, indicate the lack of relationship between adult intelligence and conformity to such school-copy values as legibility, aesthetic appearance, and neatness. In children, however, since they learn to write while under strong external pressure to conform to the school-copy model, correlations of about .30 have been found by Gesell (5) and Starch (10) between ratings of handwriting quality and general scholarship.

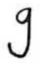
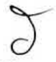
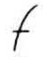
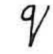
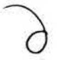
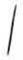
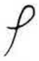
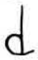
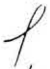
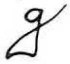
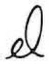
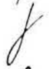
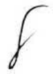
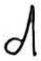
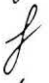
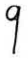
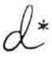
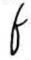
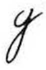
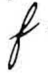
The Thorndike and Omwake studies, in having established that adherence to school-copy values in adult handwritings is not positively related to intelligence, raised the further question of what departures, if any, from the school-copy might be associated with intelligence in adult subjects. A study performed by Castelnuovo-Tedesco (3) points the way. He instructed six untrained judges to rate 100 adult handwritings for intelligence and originality, separately, and then correlated these ratings with I.Q. scores. As criteria for intelligence, the judges were told to use aesthetic spacing, handwriting speed, and the simplification (and/or) deviations of the letter forms. They were told to use as the main criteria for judging originality "the extent to which the writer of a

---

*All footnotes will be found at end of article.

TABLE I

g, d, AND f VARIANTS AND THE FREQUENCY AT WHICH THEY OCCUR AT LEAST ONCE IN A WRITING SAMPLE
(N = 150)

g		d		f	
Variant	f	Variant	f	Variant	f
	13		6		21
	19		7		21
	34		25		31
	50		66		43
	60		129		66
	74	 *	136		95
 *	143			 *	137

*School copy

given script has followed, in the process of writing a specific letter, a different-writing-pattern from that which he was probably taught in school." These correlations were all at least .50, and since 100 samples were rated it is extremely unlikely that the obtained relationships came about as a result of sampling errors.

The only other study in which judges were successful in predicting handwriting from intelligence is that reported by Binet (2). He used professional graphologists. Unfortunately, however, the handwriting criteria upon which these graphologists based their predictions have not been indicated.

In the present study, our purpose was to explore the relationship of particular deviations from the school-copy letter forms to the educational level, the I. Q. and the age of a group of female writers.

Method

Samples of the writing of 150 female clients were selected from the files of the NYU Testing and Advisement Center. An examination of them suggested three letters as worthy objects of investigation because they permitted wide graphological variations within the three handwriting zones—upper, middle and lower, without loss of identity. These letters were g, d, and f. The more predominant script variations of these letters within our sample were then noted.

In addition to the school-copy form, there were five form variants of the letter d, six form variants of the letter g, and six of the letter f. For purposes of analysis, the appearance of at least one instance of a particular letter variant in a writing sample was considered sufficient to describe that variant as present in this sample. For each variant of the letter g, d, and f, the number of samples in which it was present was determined. For instance, the school-copy g appeared at least once in 143 of the 150 writing samples. Table I presents the form variants for each of the letters and the number of Ss who used each variant at least once.

The experimenters also were struck by an additional handwriting phenomenon. Some Ss economized in a special way in writing the word "of". They would convert the upstroke normally used to round off the letter o directly into the initial upstroke of the following f. We selected this phenomenon for separate study and it is shown in Table II, which also presents the number of samples in which it was present at least once.

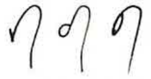
The ages, I. Q.'s and the highest grades attained by the subjects were recorded. The ages ranged from 15 to 55 and the median age was 19 years. The I. Q. scores, which were either scores on the Wechsler Bellevue Intelligence Scale or the Wechsler Adult Intelligence Scale, ranged from 80 to 145. The median I. Q. score was between 124 and 125. It was decided to use 125 as a cut-off point for various statistical tests.

As for educational level, the subjects at the lowest level had completed the eleventh grade, and the subjects at the highest level had completed more than four years of college (Table III). The subjects were divided into two groups on the basis of grade completed, one group consisted of those who had gone no further than the twelfth grade (the first two groups on Table III), and the other consisted of those who had gone to college for at least some period.

In sum, the distributions of scores on three variables, age, I. Q. and grade completed, were dichotomized for purposes of doing statistical tests. The best approximation of the median was used as the cut-off point in each case.

TABLE II

"Of" CONNECTIVE AND THE FREQUENCY AT WHICH IT OCCURS AT LEAST ONCE IN A WRITING SAMPLE
(N = 150)

Variant	f
	45

Characteristics of the Sample

A preliminary analysis of the criterion variable was made to throw light on the nature of the sample. The correlation between age and education was .42. The vast majority of subjects were still in school so that this correlation was to be expected, since educational level depends upon age among subjects who are of school age.

The correlation between age and I. Q. was .27. The older subjects in the sample were predominantly those who had continued their education beyond the period of compulsory attendance. Undoubtedly such subjects are mainly from the upper intelligence range, so there was a selective factor in favor of high I. Q. scores at the upper age levels. The correlation between education and I. Q. was .34 for the same reason.

Results and Discussion

The relationship of the d variant to age, education and I. Q. are presented in Tables IVa and IVb. Table IVa, containing the phi coefficients for the more frequently appearing d variants, shows only one significant negative relationship, a negative one between age and the third d variant. (Signif -

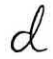
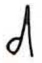
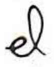
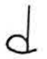
TABLE III

DISTRIBUTION OF THE HIGHEST GRADES COMPLETED BY THE 150 FEMALE SUBJECTS

Highest Grade Attained	Number of Subjects
Eleventh grade or lower	26
Twelfth grade	59
Some college	35
Four years of college or more	30

TABLE IVa

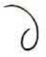
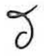
CORRELATION COEFFICIENTS OF d VARIANTS WITH EDUCATION, I. Q., AND AGE

Variant	Education	I. Q.	Age
	.12	.03	-.05
	.04	-.12	.10
	.09	.05	-.19*
	.08	.04	-.13

*Significant at .05 or above.

TABLE IVb

CHANCE PROBABILITY FOR RELATIONSHIPS BETWEEN GREEK d, MODIFIED GREEK d, AND EDUCATION, I. Q., AND AGE*

Variant	Education	I. Q.	Age
	.007	.181	.10
	.005	.356	.220

*The one-tailed level of significance is read directly from the table.

icance is considered established at the .05 level unless otherwise indicated.) Table IVb, containing the exact probability ratios for the two infrequently appearing d variants, shows each of these variants to be positively associated with educational level at better than the .01 level on a one-tailed test of significance. A one-tailed test was used because of the suggestion in the literature of the relationship between these d forms and intellectual ability (8). (The exact probability ratios for age and I.Q. showed a positive relationship between these d forms though this relationship did not reach significance.) Because of its resemblance to the Greek alphabet delta, graphologists have traditionally referred to the d variant presented first in Table IVb as the Greek-d. We shall refer to both of these variants as the Greek-d types.

The phi coefficients for the f and g variants are presented in Table V. The variants are ordered so that the one having the most positive correlation with educational level is listed first and the variant having the strongest negative correlation is listed last. Table V also shows the phi coefficients with I.Q. and age. Incidentally, the phi coefficient reflects the frequency of appearance of the variants as well as the strength of their relationships to other variables, so that comparisons should be made with caution.

Table V shows that there are three f variants which have significant positive associations with educational level, although none are significantly associated with I.Q. Among the g variants, two have significant positive associations both with educational level and I.Q. These are identified as the figure-8 and half-figure-8 g. The figure-8 g also has a significant positive association with age. One g variant is significantly and negatively associated both with age and educational level.

Considerable deviation from the school-copy model is reflected in those d, f, and g variants which were found to be associated with higher educational level, I.Q. or age. The question arises how these variants deviate from the school-copy model. We shall consider these variants, first, with respect to the relative degree of simplification which they have undergone and, second, with respect to the relative degree of motoric economy with which they are written.

Simplification is rated roughly in terms of how much graphite or ink is used to make a particular variant, as compared with the amount used in the school-copy form. Among the variants positively associated with educational level, the Greek-type d's, first and second f variants in Table V and the half-figure-8 g all involve a considerable degree of simplification. Yet, the third f in Table V is among the least simplified f variant, and the figure-8 g, which correlates more highly with educational level, I.Q. and age than any other single variant, is less simplified than the last g in Table V which is negatively associated with age and educational level. It is apparent that simplification alone cannot account for the obtained relationships.

Turning now to economy, in graphology, economy usually refers to the amount of energy and time which is involved in the writing. Hence, a letter form is judged more economical if less time elapses and less energy is utilized between the termination of the preceding letter and the beginning of the following letter. An emphasis on rightward tending strokes, an avoidance of leftward tending strokes, and a minimum of breaks and stoppages are therefore the key criteria.

The d, f, and g variants, with the exception of the second f on Table V, that were found to be significantly associated with higher educational level, I.Q. or age all show a high degree of economy. However, the second d in Table IVa, the fourth and fifth f variants in Table V, and the third and sixth g variants in Table V are also very economical forms but not significantly associated with age, I.Q. or education. Therefore, economy alone also cannot account for the relationships which were obtained.

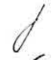
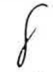
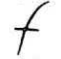
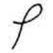
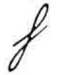
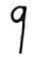
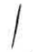
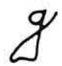
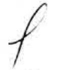
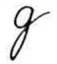
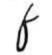
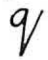
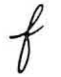
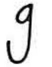
Beyond having undergone simplification and economization, those d, f, and g variants which were positively associated with higher educational level, I.Q. or age, also represent the most original deviations from the school-copy form. Considering the f variants from this viewpoint, we can see that nothing original has been evolved in any of the three variants following the school-copy one in Table V. The transformations from the school-copy have been mainly in the direction of simple omission—of either, or both upper and lower loops. In each of the three f variants which were positively associated with educational level we see the emergence of a new form relative to the school-copy model.

In the case of the figure-8 and the half-figure-8 forms, the transformation when compared with other g forms seems even greater than in the case of the f. It is interesting to note what the American graphologist, Louise Rice (8) has said concerning the figure-8 g; "People who use this formation. . . have a certain amount of intuition and instinct; they are the people who innately judge literature and art fairly well, even though not trained for such judgments. . . This g also shows innate understanding of humanity."

The upper and middle zones of the Greek-d did not differ in as specific a fashion from their school-copy counterparts as did the zonal elements of the higher f and g forms. Yet, the Greek-d types taken as Gestalten, seem to embody an aesthetic simplicity not found in the other d variants. Of the Greek-d, Louise Rice (8) has said, "It is not surprising to find stylists among writers and poets using this d exclusively, nor to find that college

TABLE V

CORRELATION COEFFICIENTS OF g AND f VARIANTS WITH EDUCATION, I. Q., AND AGE

f				g			
Variant	Education	I. Q.	Age	Variant	Education	I. Q.	Age
	. 19*	. 01	. 13		. 33**	. 23**	. 19*
	. 19*	. 16	. 10		. 23**	. 19*	. 10
	. 17*	. 13	. 08		. 05	. 05	. 13
	. 15	. 00	. 06		-. 04	. 02	-. 04
	. 14	-. 06	. 05		-. 09	-. 08	-. 06
	-. 03	-. 14	-. 15		-. 13	. 00	. 10
	-. 11	-. 09	-. 19*		-. 22**	-. 10	-. 18*

* Significant at . 05 level.
**Significant at . 01 level.

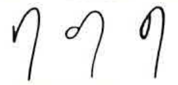
professors and scientists innately take to it. When found in ordinary writing, this formation is the 'straw in the wind' which tells of the intellectual possibilities of the writer. . . . "

Turning now to the d and f variants which were negatively associated with age, and to the g variant which was negatively associated both with age and educational level, we note that none of the three shows any evidence of originality. The f, of course, is the school-copy model; the g variant may be described as an "unfinished" school-copy g, and the d variant is a somewhat deteriorated or inept school-copy d.

Table VI reveals the "of-connective" to have a significant positive association to educational level, I. Q. and age. The direct conversion of the upstroke of the o into the initial upstroke of the f is a radically economical measure which transforms the separate letters into a unified, simplified Gestalt.

TABLE VI

CORRELATION COEFFICIENT OF "of" CONNECTIVE WITH EDUCATION, I. Q., AND AGE

Variant	Education	I. Q.	Age
	. 31*	. 23*	. 35*

*Significant at . 01 level.

Among the twenty letter-form variants which we have studied, only the figure-8 g and the half-figure-8 g were found to have a significant, positive association with higher I.Q. The "of-connective"—the only connective form which we included in this study—was also found to be associated with higher I. Q. At this point, we might wonder whether a comparably extensive study of other connective-form variants might yield a greater number of significant positive correlations with I. Q. than we have found among letter-form variants. To the extent that the need to produce a more or less legible script limits one's freedom to forge new pathways, the limitation may be greater on the individual letter forms than on the modes of combining letter forms.

Summary and Conclusions

Each of the small-letter forms selected for study were written at least once in the school-copy manner by about 90 percent of our group of 150 late adolescent and adult female Ss. In addition to the school-copy forms, we were able to classify five variants of letter d, six variants of letter f, and six variants of letter g. These variants appeared in far fewer handwritings than did the school-copy form. Also selected for separate study was a highly economical variation of the word "of".

Among the d, f, and g variants, some adhered rather closely to the school-copy form and some deviated to a much greater extent. We found the school-copy f and one d variant to be significantly associated with lower age, and one g variant to be significantly associated with both lower age and lower educational level. Both the d and f variants departed only slightly from their school-copy counterparts.

Among the seven letter-form variants which were shown to be associated with higher educational level, there were two d variants, three f variants, and two g variants. Both g variants were also associated with higher I. Q. and one, with higher age. All seven of these variants are clearly set off from the remaining eleven variants in that they combine a high degree of motoric economy with the most radical and original departures from the school-copy form. They combine efficiency with creativity and can be considered the most highly evolved of all the variants we have studied. The special case of the of-connective, which is graphologically also a highly developed variation, was also found to be associated with higher educational level, higher I. Q. and higher age.

In effect, our results suggest that the female adult who continues to write in the fashion that would have pleased her elementary school teacher is less likely to be as well educated, as bright, or as mature as the adult who has worked her way out of the school-copy rut and has evolved a more efficient and original way of writing.

The results of the present study, in revealing that there is an evolutionary trend away from the school-copy model, have led us to wonder whether the values underlying the teaching of handwriting should be revamped. Instead of attempting to cast all children in the school-copy mold, perhaps the teacher should be taught to evaluate handwriting progress in the child with respect to the efficiency and legibility of his letter forms as they appear in the context of cursive writing. With this orientation, the teacher may be less likely to make the error of diagnosing spontaneous, creative and economical deviations from the school-copy as an expression of poor progress. Why shouldn't the child who, for one reason or another, can never be a facile copyist, be encouraged to persist in whatever unique modes of writing he discovers, so long as they are no less efficient nor legible than the school copy? It would seem that an early acceptance by teachers of unique styles of writing would make writing more of a pleasure than a chore for the child who is learning to write. More important is the likelihood that many children might be spared repeated experiences of failure to write "correctly". Such experiences might be the predominant cause of the immature, inept handwritings that are seen in later years—handwritings which persist as foci of self-consciousness, embarrassment, and shame.

FOOTNOTES

* The authors wish to express their appreciation to Dr. Wallace Gobetz of the New York University Testing and Advisement Center, whose cooperation made this study possible.

** The Handwriting Institute, Inc., is a privately supported research organization formed to investigate graphological and graphomotor variables, and to collate the research of others in this area.

1. In learning the school-copy model, we are all taught, when we write words beginning with any one of these four letters, that our first stroke should be an upstroke, originating at the base line. Since this beginning stroke is completely superfluous to the essential letter form, and since it tends to reduce the efficiency and the fluency of the writing, people tend to omit it as they mature.

2. Scales of handwriting quality such as those of Ayres (1), Thorndike (12), Rosenhaus (9), are commonly used in the elementary schools to evaluate the individual child's progress in cursive writing. The scales all typically consist of a graded series of handwriting samples, from worse to better, the best and highest grade sample always being an ideal school-copy model.

REFERENCES

1. Ayres, L. P. A Scale for Measuring the Quality of Handwriting of Children (New York: Russel Sage Foundation, Division of Education, 1917).
2. Binet, A. Les Revelations de l'Ecriture d'apres un Controle Scientifique (Paris: Alcan, 1906).
3. Castelnuovo-Tedesco, P. "A Study of the Relationship Between Handwriting and Personality Variables," Genetic Psychological Monographs, XXXVII (1948), pp. 167-220.
4. Epstein, L. and Hartford, H. "Some Relationships of Beginning Strokes in Handwriting to the Human Figure Drawing Test," Perceptual Motor Skills, IX (1959), pp. 55-62.
5. Gesell, A. "Accuracy in Handwriting as Related to School Intelligence and Sex," American Journal of Psychology, XVII (1906), pp. 394-405.
6. Lewinson, T., and Zubin, J. Handwriting Analysis (New York: King's Crown Press, 1944).
7. Omwake, K. T. "The Value of Photographs and Handwriting in Estimating Intelligence," Public Personnel Studies, III (1925), pp. 2-15.
8. Rice, L. Character Reading from Handwriting (New York: Frederich A. Stokes Co., 1927).
9. Rosenhaus, M., and others. Survey Chart for Cursive Handwriting (New York: Board of Education of City of New York, 1949).
10. Starch, D. "The Measurement of Handwriting," Journal of Educational Psychology, IV (1913), pp. 445-464.
11. Thorndike, E. L. "Handwriting," Teachers College Record, XI, No. 2 (1910).
12. Thorndike, E. L. Handwriting Scale (New York: Bureau of Publications, Teachers College, Columbia University, 1910).

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 4, June 1961)

A STUDY OF PATHOLOGICAL DOUBTING AS A RESPONSE SET

VELMA I. RUST*
University of Illinois

FOR YEARS test makers have been concerned with designing tests which will measure accurately knowledge of content, specific abilities or skills, personality traits, and so on. They are interested in eliminating, or keeping to a minimum, all extraneous effects caused by guessing, or mental or attitudinal sets on the part of the subjects. Validity is kept high by giving special attention to the wording of items, their arrangement, and keying.

A few writers (3, 4) recently have raised questions regarding the relationship between evidence of knowledge of content (in the form of test scores) and a tendency to be very hesitant about accepting any statement as proven. The author became interested in studying the response set herein called pathological doubting. The definition of a pathological doubter used by Ennis (3) is the one also used here. According to him, a pathological doubter is a person who is not just cautious, but who will never accept anything as proven, a person for whom "proof", "sufficient evidence", and "warranted" are inapplicable in all conceivable situations.

The major purpose of this research was to test two hypotheses: 1. pathological doubters, chosen by the use of criteria based on their responses to selected items of the Watson-Glaser Critical Thinking Appraisal, consistently exhibit a pattern of reasoning which differs from that of the other subjects in the sample; and 2. the presence of a response set such as pathological doubting affects the apparent factor content of a test of critical thinking.

Related Research

Ennis (3) has considered the validity of the Watson-Glaser Critical Thinking Appraisal (hereafter referred to as the WG test) from a logical point of view. He asserts that the items in Parts 2, 3, and 4 of Form Bm of this test are keyed in such a way as to favor the pathological doubter. His reasoning may be illustrated by referring to Part 2 of the test, which is entitled "Recognition of Assumptions" and includes items 21 to 36 inclusive. For each item, the student is asked to decide whether a given statement "necessarily" assumes another given statement. Of the sixteen items in this part, only four are keyed "assumption necessarily made". Thus, the pathological doubter gets three-quarters of them right. Ennis then analyzes Parts 3 and 4 in the same way and shows that in the whole test the consistent pathological doubter could get up to sixty-seven out of the total of ninety-nine items correct. He, therefore, concludes that the keying so favors the pathological doubter that it is possible for him to be placed at the eighty-sixth percentile if he merely always chooses the "doubting" answer.

Frederiksen and Messick (4) have written an article in which they describe how they attempted to score set and content aspects of a test situation separately. Three tests of the report-writing type were used. The three hypotheses tested may be stated briefly as follows: a) set scores from all three tests measure the same thing reliably; b) set scores are related to personality or other set measures; and c) content scores measure the same things reliably. The investigators also tried to discover whether it was "criticalness" that the set scores measured, and whether verbal knowledge and reasoning were included among the things measured by the content scores. Frederiksen and Messick believe that if the answer to a question is not affected by knowledge, it is as if a subject merely tossed a loaded coin. They, therefore, assumed dichotomous data and derived a simple formula for set scores and another for content scores. They obtained correlation coefficients of 0. 82 when they intercorrelated content scores, 0.96 when they intercorrelated set scores, and 0. 05 when they intercorrelated content and set scores. Thus they were successful in separating content scores from set scores.

The Illinois Curriculum Program Committee (12) carried out a Critical Thinking Project in which a deliberate attempt was made to improve the critical thinking skills of a group of high school students. In order to measure differences between experimental and control groups, three tests were used: Form Bm of the WG test, A Test of Critical Thinking, Form G, prepared by the American Council on Education (hereinafter called the ACECT test), and A Test on Principles of Critical Thinking, Form F1.5 (hereinafter called the PCT test), prepared by the Illinois Critical Thinking Project Committee. The major hypothesis that knowledge of the principles of critical thinking is related to the ability to evaluate arguments and assess conclusions drawn was not confirmed. Moreover, low intercorrelations were obtained among the three critical thinking test scores. The Committee, then, began to think that the explanation for this might be in the measuring instruments themselves.

---

*The author is deeply indebted to Dr. Kenneth B. Henderson and Dr. Philip J. Runkel of the University of Illinois for their assistance with the preparation of this article.

The computations for this paper were done on Illiac, an electronic computer of the Digital Computer Laboratory of the University of Illinois.

The author (9, 10), therefore, factor analyzed the item scores for these three tests. She used a sample of 587 students consisting of all students who had attempted all items on all three tests, and showed that this method of selecting the sample did not create bias. Phi correlation coefficients were computed and the centroid method of factoring employed. Factor analyses of the three critical thinking tests yielded only one weak general factor each. In view of the fact that intelligence tests, tests of psychomotor and physical abilities, performance tests, and tests of occupational abilities have all yielded a thoroughly established general factor (within a given test), the finding in this case was rather surprising. Moreover, several investigators who have factor analyzed reasoning tests using procedures similar to that used by the author, have successfully isolated several factors (1, 2, 5, 6, 8). The author suggested several explanations for her finding, one of which was that the population lacked randomness because of the presence of too many pathological doubters. It was thought possible that instruction in the principles of critical thinking had increased the subjects' tendency to be extremely cautious. The present study was, therefore, undertaken in the hope of learning more about the trait herein called pathological doubting.

Method and Results

Item scores on the three critical thinking tests mentioned above for the same sample of 587 students were used. An attempt was made to isolate those subjects suspected of being pathological doubters. Ennis' assertion that the items in subtests 2, 3, and 4 of Form Bm of the WG test are keyed to favor the pathological doubter led the author to use the scores of the subjects on these items as the basis for dividing the 587 students into two groups, namely, pathological doubters and non-doubters. Of the 65 items in subtests 2, 3, and 4, 52 of them are keyed as "assumption not necessarily made", "conclusion does not necessarily follow", or "conclusion does not follow beyond a reasonable doubt". The author arbitrarily decided to accept getting more than 33 of these items correct as evidence of pathological doubting. The criteria used in selecting pathological doubters, therefore, were: a) having at least 34 of the 52 items keyed "assumption not made" or "conclusion does not follow" correct, and b) having no more than 19 errors among the 65 items. The second criterion was added because it was thought that students who were making many errors, and especially if they were making errors of opposite kinds, were not consistent in their thinking and, therefore, probably were confused and perhaps just guessing the answers. These criteria yielded a group of 262 pathological doubters, and a group of 325 non-doubters.

Item scores for these subjects were available for Form Bm of the WG test, Form G of the ACECT test and Form F1.5 of the PCT test. The 204 items of these three tests were divided into 21 subtests. The items of the WG test have been grouped by the test makers into five parts: Inference (20 items), Recognition of Assumptions (16 items), Deduction (25 items), Interpretation (24 items), and Evaluation of Arguments (14 items). The 52 items of the ACECT test and the 53 items of the PCT test were grouped on the basis of natural breaks into nine and seven subtests respectively. Natural breaks sometimes were in the form of asterisks placed there by the test makers, and sometimes were merely considered "natural" by the author because they came between groups of similar items or items relating to one situation. These sixteen subtests, which varied in size from 3 to 13 items, were given the short appropriate labels shown in Table IV for easy reference. Subtest scores for the groups of pathological doubters and non-doubters were obtained.

Testing Hypothesis 1. — Pathological doubters chosen using criteria based on their responses to selected items of the WG test consistently exhibit a pattern of reasoning which differs from that of other subjects in the sample.

Mean subtest scores were computed for the 262 pathological doubters and for the 325 non-doubters. The difference between the mean subtest score for the group of pathological doubters and that for the group of non-doubters, for each of the 21 subtests, was tested for significance. In each case the variance was pooled and a t-test was made using this pooled variance. Wherever the frequency distribution was not approximately normal, a χ^2 test was also made.

For each of the 21 subtests, except subtest 20, a t-test indicated that the difference between the mean subtest score for the 262 pathological doubters and that for the 325 non-doubters was significantly different from zero beyond the .01 level of significance. Differences of means were found to be significantly different from zero at the .01 level for nine of the ten χ^2 tests made. In the case of subtest 20 again, the difference was not significant. Table I summarizes these findings.

A second method was used to test this hypothesis. Answers to items 1 to 20 in subtest 1 of the WG test were scaled according to whether they gave positive or negative evidence of doubting. The scaling of these responses and the subsequent analysis were undertaken on the reasoning that people who give the doubting answer to some items (for example, those in subtests 2, 3, or 4) will give doubting answers to these items also. In subtest 1, which is called "Inference" by the test makers, a subject is required to label a statement as definitely true (T), probably true (PT), having insufficient data (ID), probably false (PF), or definitely false (F). The scale value allotted a subject for an item depended upon the keyed answer and the response given. Table II indicates how the various answers were scaled. This procedure may be illustrated by reference to item 2, which is keyed as probably true, and to a three point scale consisting of the points T or F, PT or PF, and ID. A person who is prone to accept statements as true without critically analyzing them might respond true. On the three point scale, this is one place to the left of the keyed response and indicates ready acceptance of a statement rather than a tendency to doubt. The answer, therefore, is given a scale value of -1. A response of ID is one place to the right of PT and indicates a tendency to be even more doubtful than the test makers consider warranted. Such a response was allotted a "doubting" value of +1. Other scale values were determined in a similar way. For each subject a score was allotted for each of the 20 items of subtest 1. An overall doubting index was thus obtained for each subject by adding the 20 scores. The mean doubting score for the group of 262 pathological doubters was compared with that for the group of 325 non-doubters, using a t-test.

The scaling procedure outlined in the preceding paragraph was repeated, except that in obtaining a doubting score for an item four additional scale values were also included. These are listed in the last four lines of Table II. The thinking behind this revision was that the direction of error does not matter; all that is of interest here is whether or not a subject chooses a "doubting" answer. Again, a t-test was used to test the significance of the difference of mean doubting scores of the two groups.

When the mean doubting index for the group of 262 pathological doubters (0.50) was compared with that for the group of 325 non-doubters (-0.31), a t value of 2.89 was obtained. When the revised scaling method was used, the difference between the mean doubting indices for the groups of 262 and 325 subjects (0.64 and -0.09 respectively) yielded a t value of 2.46. Thus the mean difference is significantly greater than zero beyond the .01 level of significance in the first case, and beyond the .02 level of significance in the second case.

Testing Hypothesis 2.—The presence of a response set such as pathological doubting affects the apparent factor content of a test of critical thinking.

A factor analysis of item scores for the WG test was carried out for the group of 325 non-doubters. Since the data were dichotomous, inter-item phi correlation coefficients were computed. The centroid method of factoring the correlation matrix was employed, and ones were placed in the diagonal cells of the matrix. The findings were compared with the results obtained by the author (9, 10) for the original sample of 587 which consisted of all subjects involved in the Illinois Critical Thinking Project who took all three critical thinking tests and who attempted all items on all three tests. Item scores for the PCT test were also factor analyzed using the same procedure.

As was the case in the author's original research (9, 10), no strong general factor was obtained for either the WG or the PCT test, when item scores for the 325 non-doubters were factor analyzed. In the case of the WG test the first centroid factor accounted for only 4.90 percent of the variance. For the PCT test this figure was 5.48 percent. For this reason, item scores for the ACECT test for the group of 325 non-doubters were not factor analyzed.

A factor analysis of subtest scores for the 21 subtests of the three critical thinking tests was carried out for the original sample of 587 subjects. Similar factor analyses were made for the group of 262 pathological doubters and the group of 325 non-doubters. Pearson product-moment correlation coefficients were computed. Unities were placed in the diagonal cells of the correlation matrix. Centroid factors were obtained and an attempt was made to rotate these factors to oblique simple structure using the oblimax procedure.

These factor analyses of subtest scores yielded one rather strong general centroid factor for each of the three samples. Here the first centroid factor accounted for approximately 32 percent of the variance in the case of the total sample of 587 subjects, 21 percent in the case of the 262 pathological doubters, and 22 percent in the case of the 325 non-doubters. Table III shows the variance accounted for by the first four centroid factors in the case of all three samples. Use of Guttman's "lower bound for the number of common factors" (7) indicated that an attempt should be made to rotate four factors to oblique simple structure in each case. The oblimax method of rotation was employed. This yielded quite high factor loadings, that is, high for reasoning factors. However, the

TABLE I

MEAN SUBTEST SCORES FOR THREE TESTS OF CRITICAL THINKING

Test and Subtest	Mean for 262 Pathological Doubters	Mean for 325 Non-Doubters	t*	χ^2*
Watson-Glaser				
1	12.81	10.49	6.26	
2	12.22	9.70	16.16	
3	21.45	14.99	7.87	230.00(1)**
4	17.52	14.02	8.18	
5	10.20	8.22	8.28	
ACECT				
6	5.54	4.74	7.15	154.17(3)
7	2.33	1.41	5.38	87.52(2)
8	4.37	3.09	13.66	
9	3.47	2.64	9.49	
10	8.13	6.66	7.69	50.79(2)
11	2.46	1.78	8.87	51.92(1)
12	2.20	1.33	10.33	68.25(1)
13	7.60	6.22	9.99	46.24(2)
14	4.26	3.29	6.36	45.04(3)
Principles				
15	4.38	3.75	6.70	21.57(2)
16	4.54	3.97	5.08	
17	3.42	2.79	5.27	
18	4.45	3.59	7.07	
19	7.06	5.52	9.07	
20	.89	.79	1.36	7.08(3)
21	5.41	3.98	9.77	

* All values of t and χ^2 are significant beyond the one percent level, except those for subtest 20.
**χ^2 values were computed where the distributions were not normal. Numbers of degrees of freedom are given in parentheses.

TABLE II

SCALE FOR OBTAINING A DOUBTING SCORE FOR ITEMS 1 TO 20 OF THE WATSON-GLASER TEST

		Evidence of Doubting			
		Positive		Negative	
Keyed Response*	Items	Answer Given	Scale Value	Answer Given	Scale Value
Original Method of Scaling:					
T	1, 17	PT	1		
T	1, 17	ID	2		
PT	2, 9, 16	ID	1	T	-1
ID	3, 5, 7, 8, 11, 12, 14, 18			F	-2
ID	3, 5, 7, 8, 11, 12, 14, 18			PF	-1
ID	3, 5, 7, 8, 11, 12, 14, 18			PT	-1
ID	3, 5, 7, 8, 11, 12, 14, 18			T	-2
F	4, 6, 13, 19	PF	1		
F	4, 6, 13, 19	ID	2		
PF	10, 15, 20	ID	1	F	-1
Revised Method of Scaling Contained Additions:					
T	1, 17	PF	1		
F	4, 6, 13, 19	PT	1		
PT	2, 9, 16			F	-1
PF	10, 15, 20			T	-1

*The abbreviation T stands for definitely true, PT for probably true, ID for insufficient data, PF for probably false, and F for definitely false.

TABLE III

THE VARIANCE ACCOUNTED FOR BY THE FIRST CENTROID FACTORS WHEN THE 21 SUBTEST SCORES ARE FACTOR ANALYZED

	Percent of Variance Accounted For		
Centroid Factor	Total Sample N = 587	Pathological Doubters N = 262	Non-Doubters N = 325
I	31.53	21.14	22.13
II	5.48	6.15	6.00
III	4.55	5.74	5.48
IV	4.36	5.27	4.86

oblimax solutions for the three samples were different. They are shown in Table IV.

Conclusions

The findings show the performance of the pathological doubters to be definitely and consistently superior to that of the non-doubters. They also show that those subjects who tend to choose a doubting answer in subtests 2, 3, and 4, also exhibit the same tendency when responding to items in subtest 1. There is, therefore, sufficient evidence to state that the pathological doubters do consistently exhibit a pattern of reasoning different to that of the other subjects in the original sample of 587, and thus to accept Hypothesis 1. One can also conclude that the doubters were successfully separated from the non-doubters using the author's arbitrarily chosen criteria which were based on Ennis' comments regarding the WG test.

It is particularly interesting to note that the pathological doubters consistently made higher scores than the other subjects on all subtests. Perhaps subtests 2, 3, and 4 in the WG test are not the only ones keyed to favor the pathological doubter. On the other hand, perhaps the pathological doubter's tendency to use extreme caution about accepting statements indicates that he can evaluate arguments, knows the rules of logic, recognizes unstated assumptions, and can invent and evaluate hypotheses, which are some of the skills the three critical thinking tests were designed to measure.

Subtest 20 in the PCT test seems to differ from the other subtests. It consists of three items only, and its mean item difficulty when computed for the sample of 587 was found to be only 0.28. These facts may account for the difference.

Seventy-two percent of the original sample of 587 students were in the experimental group. Seventy-nine percent of the group of 262 pathological doubters were from this group and only 66 percent of the non-doubters. Thus, there is some evidence that instruction in the principles of critical thinking increases the tendency to be overly cautious and to choose a "doubting" answer.

When item scores were factor analyzed, only slight differences were observable between the group of non-doubters and the whole sample of 587. No strong general factor was obtained in either case for either the WG or the PCT test. In contrast, when subtest scores were factor analyzed, four factors were obtained, perhaps because subtest scores are more significant and reliable than item scores. Differences in factor content, however, were apparent. Analysis of the factor loadings given in Table IV shows that the factors obtained for the three samples (the 587, the 262 pathological doubters, and the 325 non-doubters) are not sufficiently alike to be considered the same. Not even one factor common to the three samples has been found. There is, therefore, evidence here to support the hypothesis that the presence of a response set such as pathological doubting affects the apparent factor content of a test of critical thinking.

REFERENCES

1. Blotzum, W. A. "A Factorial Study of the Reasoning and Closure Factors," _Psychometrika_, XVI (1951), pp. 361-86.

TABLE IV

OBLIMAX FACTOR LOADINGS FOR SUBTESTS OF THREE TESTS OF CRITICAL THINKING

			Total Sample N = 587				Pathological Doubters N = 262				Non-Doubters N = 325			
Subtest*	Items	No. of Items	A**	B	C	D	E	F	G	H	I	J	K	L
1. Inference	1-20	20	-11	42	-26	37	101	-04	43	-25	-06	25	11	-38
2. Assumptions	21-36	16	15	79	-03	-04	64	25	09	-31	-41	15	21	22
3. Deduction	37-61	25	03	79	01	06	97	-09	37	-15	15	-13	56	01
4. Interpretation	62-85	24	02	77	-04	-10	02	57	33	21	32	63	60	-09
5. Arguments	86-99	14	00	-14	-03	72	24	-35	-11	24	-29	28	55	08
6. Direction of Evid.	1- 6	6	04	-03	02	70	-51	28	-19	67	19	-10	57	05
7. Inferences 1	7-10	4	05	76	-04	-08	70	20	80	-01	09	05	-51	99
8. Inferences 2	11-16	6	04	71	-02	05	86	01	42	-05	12	-17	57	-08
9. Generalizations	17-21	5	01	-05	01	72	-06	02	-33	43	-04	21	101	-38
10. Assumptions	22-31	10	00	07	08	65	-01	-02	31	69	-07	-07	27	43
11. Inferences 3	32-34	3	33	35	28	40	43	15	66	22	-15	-26	-02	53
12. Inferences 4	35-37	3	-03	16	46	21	38	-16	02	36	-06	-18	02	50
13. Hypotheses 1	38-46	9	-28	24	-27	40	30	-18	04	36	01	24	91	-20
14. Hypotheses 2	47-52	6	-66	03	-12	05	-07	-03	32	69	05	58	17	49
15. Definitions	1- 6	6	-36	-34	30	40	-33	24	-07	73	23	16	-34	84
16. Semantics 1	7-13	7	-75	-13	-13	08	-03	-06	29	63	37	59	66	-05
17. Fallacies	14-20	7	-73	-12	08	-11	-05	-01	-35	39	55	19	04	35
18. Semantics 2	21-27	7	-02	-02	59	16	03	04	-28	42	18	06	-37	81
19. Reasoning Rules	28-40	13	-37	38	21	-26	49	23	-05	03	46	14	08	32
20. Assumptions	41-43	3	11	-11	67	-11	-05	63	12	09	11	-48	15	-43
21. Inferences	44-53	10	-32	45	20	-24	15	45	27	32	46	19	18	32

* Subtests 1 to 5 are in the Watson-Glaser Test, subtests 6 to 14 in the ACECT test and subtests 15 to 21 in the Principles test.

**There should be two decimal places in all factor loadings. Decimal points have been omitted to save space.

2. Corter, H. M. "Factor Analysis of Some Reasoning Tests," Psychological Monographs, LXVI, No. 8 (1952), pp. 1-31.
3. Ennis, R. H. "An Appraisal of the Watson-Glaser Critical Thinking Appraisal," Journal of Educational Research, LII (1958), pp. 155-58.
4. Frederiksen, N., and Messick, S. "Response Set as a Measure of Personality," Educational and Psychological Measurements, XIX (1959), pp. 137-58.
5. Green, R. F., and others. "A Factor-Analytic Study of Reasoning Abilities," Psychometrika, XVIII (1953), pp. 135-60.
6. Guilford, J. P., and others. "A Factor-Analytic Study of Navy Reasoning Tests With the Air Force Aircrew Classification Battery," Educational and Psychological Measurements, XIV (1954), pp. 301-25.
7. Guttman, L. "Some Necessary Conditions for Common-Factor Analysis," Psychometrika, XIX (1954), pp. 149-61.
8. Matin, L., and Adkins, Dorothy C. "A Second-order Factor Analysis of Reasoning Abilities," Psychometrika, XIX (1954), pp. 71-78.
9. Rust, Velma I. Factor Analyses of Three Tests of Critical Thinking, unpublished Ph. D. dissertation, University of Illinois, 1959.
10. Rust, Velma I. "Factor Analyses of Three Tests of Critical Thinking," Journal of Experimental Education, XXIX (December 1960), pp. 177-82.
11. Rust, Velma I., and others. "A Factor-Analytic Study of Critical Thinking," Journal of Educational Research, in press.
12. Smith, B. O., and others. Teaching of Critical Thinking, to be published by University of Illinois Press.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 29, Number 4, June 1961)

GEOMETRIC AID IN TEACHING RECTILINEAR CORRELATION

JOHN GAITO
Lake Forest College
Lake Forest, Illinois

THE PURPOSE of this paper is to present a simple geometric aid for use in the teaching of rectilinear correlation.

The coefficient of correlation may be concerned with the relationship between a number of variables. In the materials to follow, the concern will be with just two variables, X and Y. A convenient means of indicating the concept of correlation is by a diagram in which the X variable is plotted on the horizontal axis and the Y variable on the vertical axis. If the points within the axes scatter in an upward direction from left to right, in a manner which indicates that X and Y tend to vary together, a positive correlation is present. If the scatter is downward from left to right, a negative correlation is present, indicating that X and Y tend to vary inversely. If there appears to be no trend present, then zero correlation is indicated.

A convenient way of teaching rectilinear correlation is to consider the scattering present as describing an ellipse. An ellipse is completely defined by its major and minor axes. The major axis is the long axis, and the minor axis is the short axis. An ellipse varies in shape with two interesting limiting cases. One limit occurs when the major and minor axes are equal; then the figure is a circle. The other extreme is when the minor axis is equal to zero and the major axis is finite, in which case we have a straight line.

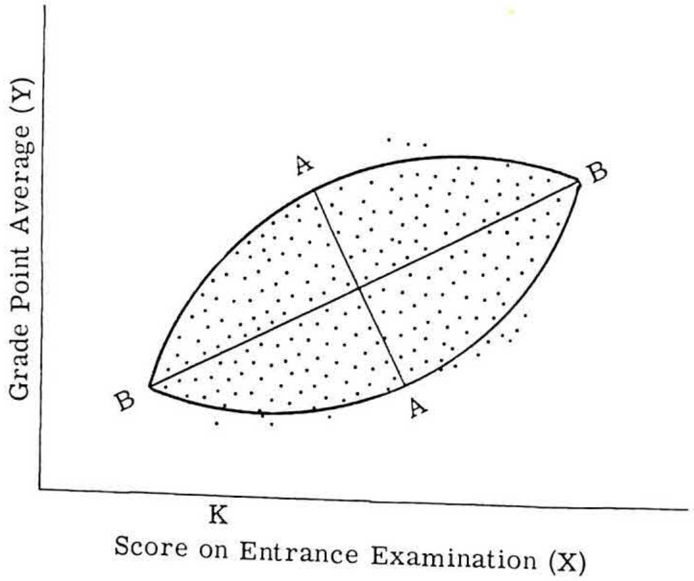
First, plot the points on the correlation chart, or scatter diagram. Thus one obtains an approximate elliptical figure by drawing a curve which encloses most points and passes through some of the outer points. Next draw a single straight line through the long axis of the figure. This line represents the trend of the scattering points. For descriptive purposes this can be handled quickly by inspection of the scatter and crudely drawing a line by hand. (A more precise method would be to draw the lines representing the regression of Y on X and X on Y, and then to use a line which is intermediate between the two to represent the trend.) This line would be the major axis of the ellipse. The minor axis would be a line drawn perpendicular to this line at its center.

Having drawn the scattering of the points to represent an ellipse, we then determine the magnitude of the relationship. This magnitude could be determined exactly by an appropriate formula available. The approximate magnitude, however, could be ascertained by looking at the ratio of the minor axis to the major axis, provided the two axes are drawn to the same scale. As the minor axis goes to zero (or the major axis gets very large relative to the minor axis), this ratio will also go to zero, and the correlation coefficient will vary toward one. When the minor axis equals zero and the major axis is finite, the ratio also equals zero, the coefficient is one, no spread is present, and we have a single straight line—with perfect correlation. (An exception occúrs when the slope of the line representing the major axis equals zero. In this case, the Y variable equals a constant value for every X, and the line is parallel to the X axis. Thus, there is no change in Y with a change in X; Y remains the same. In this case, even though the minor axis is zero and the ratio is zero, the correlation coefficient would be zero because no relationship is present.) On the other hand, as the minor axis increases, with the major axis relatively fixed, the ratio increases and the coefficient decreases. When the minor axis is equal to the major axis, the ratio is one, the coefficient is zero, the spread is maximal, and we have a circle with no relationship present.

This geometrical interpretation has value beyond the elementary teaching of correlation. For example, it is found useful in looking at the effects of range of ability in the sample on the size of the coefficient. Let us use an example in which 500 Ss have scores on a college entrance examination, all are accepted, grade-point averages are available later for each, and the relationship between examination scores and grade-point average is .60. If one were to plot these data on a correlation chart it would be found that a ratio of approximately 2/5 would occur. The ratio might be greater or less, but the magnitude of the ratio is not important for this example. Only the relative aspects are of consequence. Thus, let us assume that 2/5 is the ratio. This situation is shown in Figure 1. Now if, instead of allowing all 500 applicants to enter school, we had used a cutting score on the examination, the diagram would be

FIGURE 1

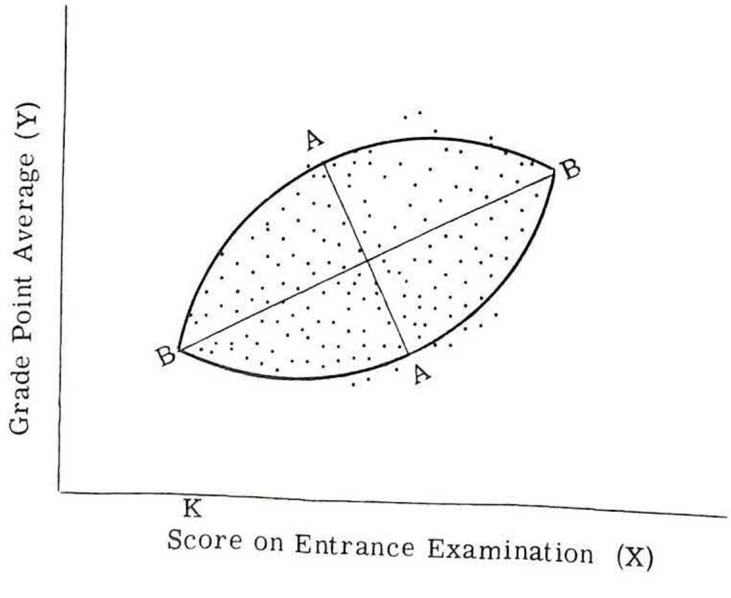
A HYPOTHETICAL EXAMPLE IN WHICH ALL APPLICANTS ARE ACCEPTED AT A COLLEGE



Note: The relationship indicated is .60. The ratio of the minor axis (A) to the major axis (B) is assumed to be 2/5. Point K represents the cutting score which is used in Figure 2.

FIGURE 2

THE RELATIONSHIP INDICATED WHEN A NUMBER OF APPLICANTS TOWARD THE LOWER END OF THE X VARIABLE, BELOW K, HAVE BEEN ELIMINATED



changed as in Figure 2. Note that the minor axis has been shifted from its position in Figure 1 so as to be perpendicular to the major axis at the latter's center; however, the minor axis is approximately the same in magnitude as it was in Figure 1. The major axis has decreased in size so that the ratio now is approximately 2/4. This increase in ratio would indicate a lesser degree of relationship than is present in Figure 1. Such a procedure easily illustrates the effect of composition of the sample (as indicated by the range of the X variable) on the magnitude of the relationship, with a greater degree of relationship present when the sample is heterogeneous than when the sample is homogeneous.

We have not developed an exact mathematical relationship between the size of the ratio and the size of the correlation coefficient. Intuitively, however, we assume that there is such a relationship, so that one could indicate the size of the coefficient by knowing the size of the ratio. This geometrical interpretation of correlation, using an elliptical ratio, should be useful for teaching purposes even if it is merely of an ordinal nature.